UNIL | Université de Lausanne

Faculté de Biologie et de Médecine

Département d'Écologie et Évolution

SOCIOGENETICS OF FIRE ANTS

Thèse de doctorat ès sciences de la vie (PhD)
présentée à la Faculté de biologie et de médecine de l'Université de Lausanne par

YANNICK WURM

ingénieur en BioSciences (filière Bioinformatique et Modélisation) de
l'Institut National de Sciences Appliqués de Lyon (France)

Jury

Prof. Christian Fankhauser, Président & rapporteur
Prof. Laurent Keller, Directeur de thèse
Prof. Laurent Excoffier, Expert
Prof. Henrik Kaessmann, Expert
Prof. Evgeny Zdobnov, Expert

Lausanne, 2009

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président	Monsieur	Prof.	Christian **Fankhauser**
Directeur de thèse	Monsieur	Prof.	Laurent **Keller**
Experts	Monsieur	Prof.	Laurent **Excoffier**
	Monsieur	Prof.	Henrik **Kaessmann**
	Monsieur	Dr	Evgeny **Zdobnov**

le Conseil de Faculté autorise l'impression de la thèse de

Monsieur Yannick Wurm

Ingénieur INSA Lyon

intitulée

Sociogenetics of fire ants

Lausanne, le 18 décembre 2009

pour Le Doyen
de la Faculté de Biologie et de Médecine

Prof. Christian Fankhauser

RÉSUMÉ GRAND PUBLIC

La complexité des sociétés d'insectes (telles que les abeilles, les termites ou les fourmis) a depuis longtemps fasciné l'Homme. Depuis le début du XIXème siècle, de nombreux travaux observationnels, comportementaux et théoriques leur on été consacrés afin de mieux les décrire et comprendre. L'avènement de la biologie moléculaire à la fin du XXème siècle a offert de nouveaux outils scientifiques pour identifier et étudier les gènes et molécules impliqués dans le développement et le comportement des êtres vivants. Alors que la majorité de ces études s'est focalisée sur des organismes de laboratoire tel que la mouche ou les nématodes, l'utilisation de ces outils est restée marginale jusqu'à présent dans l'étude des sociétés d'insectes.

Lors de ma thèse, j'ai développé des outils moléculaires permettant de déterminer le niveau d'activité de 10,000 gènes chez la fourmi de feu, *Solenopsis invicta*, ainsi qu'une base de données et un portail en ligne regroupant les informations relatives à l'étude génétique des fourmis: Fourmidable. J'ai ensuite utilisé ces outils dans le cadre d'une étude comportementale chez la fourmis *S. invicta*. Dans les sociétés d'insectes, une hiérarchie peut déterminer le statut reproducteur des individus. Suite à la mort d'un dominant, les subordonnés entrent en compétition en vue d'améliorer leur statut. Un tel phénomène se produit au sein des colonies de *S. invicta* contenant une unique reine mère, des milliers d'ouvrières et des centaines de reines vierges ailées. A la mort de la reine mère, un grand nombre de reines vierges tentent de la remplacer en arrachant leurs ailes et en activant leurs organes reproducteurs plutôt que de partir en vol nuptial. Ces tentatives sont le plus souvent arrêtées par les ouvrières qui exécutent la plupart de ces reines sur la base de signaux olfactifs produits lors de l'activation des organes reproducteurs. Afin de mieux comprendre les mécanismes moléculaires impliqués, j'ai étudié l'activité de gènes au sein des reines au début de ce processus. J'ai ainsi déterminé que des gènes impliqués dans communication olfactive, le développement des organes reproducteurs et la métabolisation de l'hormone juvénile sont activés à ce moment là. La vitesse à laquelle les reines perdent leurs ailes ainsi que les niveaux d'expression de gènes sont ensuite liés à leur probabilité de survie.

ABSTRACT

Honeybees, termites and ants occupy the "pinnacle of social evolution" with societies of a complexity that rivals our own. Humans have long been fascinated by social insects, but studying them has been mostly limited to observational and behavioral experiments. The advent of molecular biology first made it possible to investigate the molecular-genetic basis of development in model systems such as the fruit fly *Drosophila melanogaster* or the roundworm *Caenorhabditis elegans* and subsequently their behavior. Molecular and genomic tools are now becoming available for the study of social insects as well.

To permit genomic research on the fire ant, *Solenopsis invicta*, we developed a cDNA microarray that can simultaneously determine the expression levels of approximately 10,000 genes. These genes were assembled and bioinformatically annotated using custom pipelines. The obtained data formed the cornerstones for Fourmidable, a web portal centralizing sequence, gene annotation and gene expression data as well as laboratory protocols for research on ants.

In many animals living in groups the reproductive status of individuals is determined by their social status. In species with social hierarchies, the death of dominant individuals typically upheaves the social hierarchy and provides an opportunity for subordinate individuals to improve their social status. Such a phenomenon occurs in the monogyne form of *S. invicta*, where colonies typically contain a single wingless reproductive queen, thousands of workers and hundreds of winged non-reproductive virgin queens. Upon the death of the mother queen, many virgin queens shed their wings and initiate reproductive development instead of departing on a mating flight. Workers progressively execute almost all of them over the following weeks. The workers base their collective decision on pheromonal cues associated with the onset of reproductive development of the virgin queens which occurs after orphaning. We used the aforementioned tools to determine that genes putatively involved in processes including olfactory signaling, reproductive development and Juvenile Hormone metabolism are differentially expressed at the onset of competition. Additionally, we found that queens that initiate reproductive development faster and, to a certain extent, shed their wings faster after orphaning are more likely to become replacement queens. These results provide candidate genes that are putatively linked to competition outcome.

To determine the extent to which specific genes affect different aspects of life in ant colonies, functional tests such as gene activation and silencing will still be required. We conclude by discussing some of the challenges and opportunities for molecular-genetic research on ants.

RÉSUMÉ

Les sociétés d'abeilles, de termites et de fourmis sont d'une complexité proche de celle de la nôtre et ont depuis longtemps fasciné l'Homme. Cependant, leur étude était jusqu'à présent limitée aux observations et expériences comportementales. L'avènement de la biologie moléculaire a d'abord rendu possible l'étude moléculaire et génétique du développement d'organismes modèles tels que la mouche *Drosophila melanogaster* ou le nématode *Caenorhabditis elegans*, puis dans un second temps de leur comportement. De telles études deviennent désormais possibles pour les insectes sociaux.

Nous avons développé une puce à ADN permettant de déterminer simultanément les niveaux d'expression de 10,000 gènes de la fourmi de feu, *Solenopsis invicta*. Ces gènes ont été séquencés puis assemblés et annotés à l'aide de pipelines que nous avons développés. En se basant sur les informations obtenues, nous avons créé un portail web, Fourmidable. Ce portail vise à centraliser toutes les informations de séquence, d'annotation et d'expression de gènes, ainsi que les protocoles de laboratoire utilisés pour la recherche sur les fourmis.

Par la suite, nous avons utilisé les outils développés pour étudier un aspect particulier de *S. invicta*. Chez les animaux grégaires, une hiérarchie sociale peut déterminer le statut reproducteur des individus. Suite à la mort d'un individu dominant, les individus subordonnés peuvent entrer en compétition en vue d'améliorer leur statut. Un tel phénomène se produit au sein des colonies monogynes de *S. invicta*, qui contiennent habituellement une unique reine mère, des milliers d'ouvrières et des centaines de reines vierges ailées. Suite à la mort de la reine mère, dominante, un grand nombre de reines vierges, subordonnées, perdent leurs ailes et activent leurs organes reproducteurs au lieu de partir en vol nuptial. Au cours des semaines suivantes, les ouvrières exécutent la plupart de ces reines sur la base de signaux olfactifs produits lors de l'activation des organes reproducteurs. Afin de mieux comprendre les mécanismes moléculaires impliqués, nous avons étudié l'expression de gènes au début de cette compétition. Nous avons identifié 297 gènes différemment exprimés, dont l'annotation indique qu'ils seraient impliqués dans des processus biologiques dont la communication olfactive, le développement des organes reproducteurs et la métabolisation de l'hormone juvénile. Par la

suite, nous avons déterminé que la vitesse à laquelle les reines perdent leurs ailes en début de compétition ainsi que les niveaux d'expression de gènes sont corrélés à la probabilité de survie des reines.

Nous concluons en discutant des opportunités offertes par la recherche génétique sur les fourmis ainsi que les défis qu'elle devra surmonter.

ACKNOWLEDGMENTS

Au cours de ces cinq années de thèse et de vie à Lausanne, j'ai eu la chance de cotôyer des personnes qui m'ont inspirées, aidées et soutenues. C'est grâce à elles que Lausanne est devenu mon chez moi où je me sens bien. Voici une petite tentative de les reconnaître.

J'aimerais tout d'abord commencer par remercier mon directeur de thèse **Laurent Keller**. Ce fut un honneur de pouvoir travailler avec toi. Je te remercie de m'avoir tant appris sur l'évolution, l'écriture et la rigueur scientifique, pour avoir été un leader charismatique et efficace, pour m'avoir permis de grandir en me laissant gérer librement mon tempset mes projets. Merci aussi de m'avoir montré que d'être un scientifique du plus haut niveau n'empêche pas de pouvoir prendre plaisir à la vie et faire la fête.

Ensuite, un très grand merci à **John Wang**. You have been a meticulate, enthusiastic and caring mentor. Your understanding of all genomic and evoltuionary things is tremendous and I am infinitely grateful for your patience and guidance. Science needs more people like you.

Thanks to my committee members, **Evgeny Zdobnov**, **Henrik Kaessmann, Laurent Excoffier** and **Christian Fankhauser** for taking the time to look through my work and for providing constructive criticism and stimulating discussions.

Merci à mes collègues de bureau **Rob Hammond**, **Joël Meunier**, **Valou Vogel**, **Christophe Lucas**, **Hélène Niculita** pour avoir été mes confidants, pour avoir partagé de longues discussions stimulantes, pour les moments délirants, pour avoir toléré mon bordel, et globalement pour avoir contribué à mon plaisir de travailler au DEE. À **Christine LaMendola**, **Michael Nicolas**, **Miguel Corona** et **DeWayne Shoemaker** pour leur enthousiasme et les collaborations fructueuses. À **Sévérine Büchel** pour son énergie et ses mots de suisse-allemand tout au long du Master. À **Steffi Jemielity** pour m'avoir convaincu qu'on peut avoir des "ambitions normales" tout en travaillant sur les fourmis ;) À ceux qui ont tant aidé pour la programmation (**Fred Ricci**) pour nourir les fourmis (**Jérôme Notari**, **Jézaëlle Rufener**, **Nono**, **Yann Hautier**, **Aurélie Righetti**) et au labo (**Nathalie Tzaud**, **Françoise Dolivo**).

À **Marc Robinson-Rechavi**, **Julien Roux**, **Sara Mitri**, **Sébastien Morretti**, **Romain Studer**, **Steffen Wischmann** et **Nicolas Salamin** pour leur précieux conseils statistiques et computationels et les moments de geekiness partagés.

À **Franck Chalard**, **France Pham**, **Fély Jacquiery**, **Corine Bolle**, **Virginie Cantamessa** et **Blaise Pavillard** dont l'excellent travail nous protège des aléas administratifs et techniques de la géstion du département.

À **Ioannis Xenarios**, **Laurent Falquet**, **Keith Harshman**, **Frédéric Schütz**, **Darlene Goldstein**, **Christian Iseli**, **Volker Flegel**, **Manuel Bueno** et aux staffs de **Vital-IT** et du **DAFL** pour leurs savants conseils, leur disponibilité, l'accès à leurs ressources et l'enthousisasme qu'ils ont toujours montrés pour les nouveaux projets.

Aux autres collègues du groupe fourmis, **Greg Castella**, **Dietrich Gotzek**, **Mingkwan Nipitwattanaphon**, **Lino Ometto**, **Tanja Schwander**, **Mischa Djikstra**, **Catherine Berney**, **Danielle Mersch**,

Romain Libbrecht, **Thibault Lengronne**, **Joel & Karen Parker**, **Michel Chapuisat**, **Anabelle Reber** et ceux qui étaient aussi de passage pour les soirées festives et les discussions stimulantes.

Aux membres du SIB PhD School/Training network pour les cours et les discussions stimulantes ainsi que les belles soirées.

Aux membres du EGG group, de Toastmasters, de l'équipe de foot du DEE, aux FIPEurs, aux amateurs de Karaoké, à l'équipe et aux étudiants de Bulgarie des TPs de stats, de bioinfo et de prog, aux frouz du DEE, aux apéros, aux grillades, aux équipes des ESEB, IUSSI et Congrès d'Entomologie (**Lucie**, **Alan**, **Seb**, **Hélène**, **Christian**, **Aline**, **Luc!**, **Véro**, **Daniel**, **Lori**, **Walter**, **Beat**, **Alain**, **Yuko!**, **Jon**, **Reto**, **Daniel2**, **Christine**, **Charlène** et ceux que j'ai oubliés) pour tous les bons moments.

Aux étudiants qui m'ont appris le plaisir d'enseigner.

Beaucoup d'entre vous me sont devenus de proches amis. Et m'ont apporté l'équilibre tant nécessaire pour réussir une thèse. Merci aussi à ceux que je n'ai pas encore mentionnés mais qui ont aussi pris une grande place dans mon coeur (**Karen**, **Cianci**, **Godefroy**, **Nando**, **Grigoris**, **Subhash**, **Sebbyseb**...).

À mes musciens des Terracemen, de CrossOver, du Convulsif, de Lapointe, de l'atelier 1D et d'ailleurs, aux Karatékas de Suisse-Shotokan Lausanne.

À mes amis et ma famille lointaine, au meilleur frangin et à la meilleure frangine du monde, à mes parents qui m'ont donné tout pour réussir et à ma chère et tendre Tia. Vous avez partagé tant d'amour et m'avez soutenus avec tellement de patience. I love you guys so much!

Cette thèse est dédiée à la mémoire d'Adèle Fraboulet et de Gisela Wurm-Sieber.

CONTENTS

Part I

INTRODUCTION TO SOCIAL INSECT GENOMICS

1

GENERAL INTRODUCTION

by Yannick Wurm

1.1 EVOLUTION OF SOCIAL LIFE

New ecological niches were created and successfully occupied when multi-cellularity evolved. Similarly, the transitions from solitary to social life created new ecological opportunities and led to the ubiquitousness of social organisms, which include primates and social insects such as ants, wasps, bees and termites (Maynard Smith & Szathmáry 1995). A captivating aspect of social insects is the high behavioral, physiological and morphological specialization found between castes. For example, social insect queens can lay up to thousands of eggs every day while workers have inactive or even completely lack reproductive organs (Wilson 1971).

The existence of non-reproducing workers was "*one special difficulty, which at first appeared to me insuperable, and actually fatal to the whole theory [of natural selection]*" remarked Charles Darwin in *The Origin of Species* (1859). Indeed, not reproducing should be an evolutionary dead end because natural selection should favor the individuals with the highest reproductive success. However, Darwin (1859) also clairvoyantly noted that *"This difficulty [...] disappears [...] when it is remembered that selection may be applied to the family"*.

In fact, natural selection is now widely accepted to act at the level of individual genes (Hamilton 1964; Dawkins 1976; Bourke & Franks 1995), and is explained mathematically using the kin selection framework (discussed in West *et al.* 2007; Keller 1999). The basic logic of kin selection is that a gene can increase its transmission rate from one generation to the next either by directly increasing the fitness of its bearer, or by increasing the fitness of relatives who share copies (identical by descent) of the gene. This implies that altruistic behavior will be favored

in groups of close relatives as found in colonies of social insects (Hamilton 1964; Queller & Strassmann 1988) (Hamilton 1964; Wilson 1971; Bourke & Franks 1995; Langer *et al.* 2004; Passera & Aron 2005; Strassmann 2006). The evolution of reproductive division of labor in ants, honey bees and termites between 120 and 180 million years ago (Bourke & Franks 1995; Hedges & Kuma 2009) subsequently provided the opportunity for the evolution of morphological castes and complex social structures.

1.2 DISSECTING THE MOLECULAR BASIS OF SOCIAL LIFE

Life in societies implies the existence of mechanisms for task coordination, decision making and conflict resolution. Despite extensive studies on the behavior and ecology of insect societies as well as of the evolutionary pressures in and around them, still little is known about the molecular-genetic basis of their development and behavior. Tools such as genetic mapping, gene expression quantification and gene silencing and activation have been used to dissect *Drosophila melanogaster* development (White *et al.* 1999) and behaviors including aggression (Dierick & Greenspan 2006), courtship (Manoli *et al.* 2005) and foraging (Osborne *et al.* 1997). They can also help us to examine insect societies in molecular terms.

First steps in honey bees

Researchers began investigating the lives of social insects at a genetic level in the late 1990s. Such "sociogenetic" studies were kick-started by the construction of gene expression microarrays that could simultaneously determine the activity levels of thousands of genes in single honey bees (Whitfield *et al.* 2002). A few years later, molecular studies on honey bees received a further boost via the efforts put into the honey bee genome project (The Honey Bee Genome Sequencing Consortium 2006) (reviewed in Chapter 2).

An elegant series of experiments examined the link between gene activity and division of labor in workers (Ben-Shahar *et al.* 2002; Whitfield *et al.* 2003, 2006). Charles Whitfield and colleagues used the new tools to contrast gene expression between the brains of nursing and

foraging honey bee workers, controlling for age. Several genes and pathways had significantly different activity levels between nurses and foragers. Ectopic activation of some of these pathways induced foragers to nurse, thus confirming the roles of these pathways in the division of labor (Ben-Shahar *et al.* 2002; Whitfield *et al.* 2006). Similar approaches have been used to examine other aspects of honey bee biology. For example, pheromones produced by the queen and by the brood affect gene worker behavior and brain gene expression (respectively Grozinger *et al.* 2003; Alaux *et al.* 2009). Genes have also been identified that are differentially activated between both reproductive and non-reproductive queens and workers (worker reproduction may occur for example if the queen is removed; Kocher *et al.* 2008; Grozinger *et al.* 2007). Finally, much focus has been on the genes that determine whether a female egg will develop into a worker or a queen (Wheeler *et al.* 2006; Cristino *et al.* 2006; Grozinger *et al.* 2007; Elango *et al.* 2009; Patel *et al.* 2007).

These studies in honey bees provided major insight into the genetic basis of social life. However, honey bee colonies harbor only a single reproductive queen, while a few to thousands of queens cohabit in many ant species (Wilson 1971). Furthermore, the diversity of ant lifestyles are beyond parallel as they include trap building ants (Dejean *et al.* 2005), slave-making ants (Lenoir *et al.* 2001), nomadic army ants (Kronauer 2008) and ants that invented farming 50 million years ago (Schultz & Brady 2008). Finally, the morphological adaptations of ant castes are far more extreme than in bees. Indeed, task specialization among workers is mainly behavioral in honeybees (Robinson 1992; Wilson 1971), while many ant species have extreme morphological adaptations to their specialized tasks (Anderson & McShea 2001). Examples of this include the almost 100-fold dry weight range among *Eciton burchelli* army ant workers (Franks 1985; Franks *et al.* 1999), the *Myrmecocystus mexicanus* honey pot ants which function as food storage silos (Hölldobler & Wilson 1990) and the nest entrance plugging heads of large *Camponotus anderseni* workers that live inside mangrove roots which are submerged at high tide (Nielsen 2000). The great complexity and diversity of ant social structures, lifestyles and adaptations provide many opportunities for investigating fundamental questions about the genetic basis of social life.

Past findings in ants

The aim of most molecular-level research in ants has been to understand the genetic structure and relatedness within and between populations and colonies. This approach fortuitously provided surprising insight into different aspects of ant biology.

The first example of this concerns the remarkable complexity of the reproductive systems of ants. Sexual reproduction produces sterile workers in *Wasmannia auropunctata*. However, mother queens make clones of themselves to produce new queens. This would normally imply that maleness is an evolutionary dead end. That is not the case: *W. auropunctata* males reproduce clonally by using maternal egg cells that contain no maternal DNA (Fournier *et al.* 2005). Discoveries of such unconventional reproductive systems are ironically becoming common-place in ants (Pearcy *et al.* 2004; Fournier *et al.* 2005; Schwander & Keller 2008; Dobata *et al.* 2009; Keller 2007; Heinze 2008).

Another fortuitous discovery ten years ago was that a single genetic locus can regulate social structure. *Solenopsis invicta* colonies in which at least 10% of workers carry the *b* allele accept queens that carry that allele and execute those that do not. However, when less than 5% of colony workers carry the *b* allele, only a single queen lacking this allele is accepted (Keller & Ross 1998; Ross & Keller 1998; Gotzek & Ross 2007). This is a rare example of group genotypic composition determining individual behavior.

Recently, experiments were conducted with the specific aim of identifying genes that were differentially expressed between *Camponotus festinatus* worker larvae and adults (Goodisman *et al.* 2005), between adult *Lasius niger* queens and workers (Gräff *et al.* 2007), and after mating in *Solenopsis invicta* queens (Tian *et al.* 2004). These studies had no *a priori* expectation of which genes should be found.

Other studies are based on the idea that homologs of molecules that have been studied in other species are good candidates for playing similar roles in ants. This indeed is the case for Juvenile Hormone which plays a role in the onset of reproduction in many insects (Chapman 1988) as well as in queens of *S. invicta* (Troisi & Riddiford 1974; Kearney *et al.* 1977; Vargo & Laurel 1994; Brent & Vargo 2003; Burns *et al.* 2007). The most interesting candidate-gene

based studies focused on the *foraging* gene, the alleles of which determine the extent to which *Drosophila melanogaster* larvae actively search for food (Sokolowski 1980). The homologous gene appears to play a role in ant behavior as well. Indeed, *foraging* expression level is correlated to division of labor in *Pogonomyrmex barbatus* harvester ants (Ingram *et al.* 2005); and an elegant series of experiments showed that *foraging* activation induces *Pheidole pallidula* workers to shift from foraging to defense behavior (Lucas & Sokolowski 2009).

The aforementioned studies provided valuable insight into the genetics of social behavior in ants. However, they focused on only very small numbers of genes. Much work remains to truly understand the roles of candidate genes that putatively regulate aspects of development, physiology and behavior linked to sociality. In particular, that individual genes play specific roles can only be demonstrated by laborious follow up studies using methods that include gene activation and silencing. Candidate genes for which one can justify such efforts are difficult to find for the many aspects of social life that are unique to ants. There was thus a pressing need to develop versatile tools for identifying candidate genes in ants.

1.3 AIMS AND OUTLINE OF THIS THESIS

My aim during this PhD was to make a first attempt at identifying genes involved in social behavior in ants. When I started the PhD, several fascinating molecular studies had already been conducted in bees. These were further strengthened by the efforts of the honey bee genome project (reviewed in **Chapter 2**). However, only few molecular studies had been performed on ants and molecular tools were limited. A substantial part of my time was therefore dedicated to creating the tools necessary for genome-wide studies in ants.

Development of genomics tools for ants

Construction by John Wang and Stephanie Jemielity of a microarray carrying some 20,000 cDNA clones was well under way by the time I joined the Keller lab. My computational contribution to the project included annotation of the approximately 10,000 sequenced genes

and the identification of 23 putatively hymenopteran-specific genes (**Chapter 4**). Additionally, the database deployed to house the ant sequence data, Fourmidable, required revision because it was based on software and hardware that were becoming obsolete. We thus completely rewrote the assembly pipeline, improved the gene annotation methods, added support for microarray data and generally improved Fourmidable's user-friendliness (**Chapter 3**). Once the microarray and gene annotations were in place, we used them to examine a unique aspect of life in fire ant colonies.

Genomic investigation of a competition for reproductive dominance in fire ants

A central issue in research on social life is to understand how reproductive dominance hierarchies are maintained and how individuals react to and compete during opportunities for social ascension. We conducted two studies that should shed light on the molecular basis of such a situation.

We examined the reaction to loss of the dominant individual on subordinate individuals in the single-queen form of the fire ant, *Solenopsis invicta* (**Chapter 5**). Virgin queens shed their wings and initiate reproductive development in lieu of departing on a mating flight upon the death of the dominant queen. Workers progressively execute almost all of them over the following weeks based on pheromonal cues produced by the young queens after orphaning. We compared gene expression levels in virgin queens before and after death of the dominant queen using the microarrays from **Chapter 4**. We identified 297 genes that were consistently differentially expressed after orphaning. The putative functions of these genes indicate that they are involved in biological processes including protein transport, Juvenile Hormone metabolism and the onset of reproductive development.

We used the speed at which virgin queens from different lineages shed their wings after orphaning as a proxy for the speed at which they responded to orphaning in **Chapter 6**. This information was used to set up competitions by placing queens from pairs of lineages that were faster and slower at shedding their wings with unrelated workers. We found that queens that were faster at shedding their wings were more likely to survive the competitions. Additionally,

we found that queens that shed their wings faster have higher mitochondrial activity, and that queens that are most likely to win competitions have higher mitochondrial activity and higher activity of genes related to organ development and cell differentiation.

Finally, on **page 107** I discuss additional projects that are underway to identify genes that may be responsible for developmental differences between castes as well as between social structures, as well as some of the exciting opportunities created by applying molecular technologies to research on ants.

REFERENCES

References for this chapter are found on page 113.

2

REVIEW OF HONEY BEE GENOME EFFORT.

by Yannick Wurm, John Wang and Laurent Keller

ABSTRACT

Honeybees, termites and ants occupy the 'pinnacle of social evolution' with societies of a complexity that rivals our own. The sequencing of the honeybee genome will provide a strong foundation for studying the genetic basis of complex social behavior.

This review was published in *Current Biology* (2007) 17:2, R51-R53 using the title *Behavioral Genomics: A, Bee, C, G, T...*
DOI: 10.1016/j.cub.2006.12.019

Go to the ant, thou sluggard; consider her ways, and be wise

Proverbs 6:6

This well-known Bible verse appropriately illustrates the long fascination of human beings for the complexity of social insect colonies and the industrious nature of their workers. The major organizing principle of ant, bee and termite societies is reproductive division of labor whereby one or a few individuals, the queens, specialize in reproduction while the others, the workers, participate in cooperative tasks such as building the nest, collecting food, rearing the young and defending the colony. This social organization provides numerous advantages and is the basis for the tremendous ecological success of social insects [1]. The sequencing of the honeybee genome [2] is an exciting step towards uncovering the molecular events underlying the evolution of altruism and complex behaviors.

A genome sequence, like a honeybee queen, is useless if it is not accompanied by the assiduous labor of a large number of cooperative workers. For this reason, an industrious swarm of scientists has collaborated to conduct detailed analyses and comparisons of the honeybee genome with those of model organisms, in particular the fruitfly *Drosophila melanogaster*. These analyses, reported in no fewer than 40 companion papers published in Science, PNAS and special issues of Genome Research and Insect Molecular Biology, have revealed many interesting features associated with the unusual social biology of the honeybee.

To date, most of our understanding of behavioral genetics has come from studies in model organisms such as *D. melanogaster* and *C. elegans*. In only a handful of cases, however, have genetic variants been shown to be responsible for behavioral differences observed under natural conditions [3,4]. An interesting feature of the honeybee is that workers change tasks as they age. They typically remain in the nest when young and then switch to dangerous tasks outside the nest only when they are older. This switch can be manipulated by altering colony needs or by treatment with chemicals that cause precocious foraging. Capitalizing on this, Whitfield et al. [5] conducted a series of clever experiments to separate the effects of worker age, genotype, environment and experience on gene expression. They found that the transition from hive work to foraging is accompanied by a robust molecular signature with many genes sharing the same expression pattern across the conducted experiments. Examination of these genes revealed shared cis-regulatory promoter elements that may be responsible for their co-regulation [6]. This is a big step towards clarifying the regulatory cascades governing the networks of behavioral genes. Future investigations into the neuronal circuitry of bees and its modulation should also be facilitated by the bioinformatic and proteomic identification of 36 prohormones coding for more than 200 neuropeptides [7].

Comparison of the honeybee genome to that of other insects revealed a number of interesting differences. Not a complete surprise was the identification of nine new genes linked to the production of royal jelly, which workers feed to the queen and larvae. These genes apparently evolved from a single progenitor gene which encodes a member of the ancient Yellow protein family [8]. Similarly, the striking expansion of the odorant receptor family in honeybees (170 genes) relative to *D. melanogaster* (62) and the mosquito *Anopheles gambiae* (79) [9] makes

sense given the prime role of pheromones in communication and the need of workers to discriminate between diverse floral odors. These discoveries should help elucidate some of the bees' fascinating skills, which include precise memory of space and odors as well as the abstract modeling and linguistic abilities shown by the 'waggle-dance'.

A more surprising finding was that honeybees have only half as many immune defense genes as *D. melanogaster* or *A. gambiae*. Many authors have suggested that colonies of social insects should be under particularly strong pathogenic pressure because numerous highly related individuals live in close quarters. Although not frequently recognized, however, bees spend most of their lives in a protected colony environment, while flies and mosquitoes grow up in rotting food or stagnant water. Moreover, the food provided to bee larvae has already been processed by adults and thus is less likely to contain pathogens. Likewise, the risk of poisoning or infection may be low when foraging nectar and pollen from flowers with which a mutualism has evolved [10]. The close association between bees and plants, and the lack of incentive for plants to produce toxic nectar or pollen may actually also explain the a priori surprising finding that honeybees have far fewer gustatory receptors than D. melanogaster and A. gambiae [9]. Another possible explanation for why bees have fewer immune defense genes is that they display social behaviors such as extensive grooming and 'social fever' that may effectively combat infections [10].

The honeybee genome has provided several interesting revelations concerning the most unusual characteristic of social insects: their ability to produce very different phenotypes from the same genotype as a result of the alternate developmental programs followed by queens and workers. Indeed, in several ant species, queens can differ dramatically in size and morphology from workers, yet almost nothing is known about the epigenetic factors underlying the developmental switch responsible for these differences. Of particular interest was the discovery of 65 microRNAs, including some that show caste-specific expression patterns during development [2]. This raises the exciting possibility that microRNAs are involved in caste determination via differential gene expression between queens and workers.

The finding that the honeybee genome harbors genes encoding a complete set of methyltransferases, the highest known eukaryotic CpG content, and evidence for CpG methylation of protein coding genes is also of great interest, given that methylation of CpGs represses transcription in mammals. Interestingly, in contrast to mammals, in the honeybees DNA methylation was detected predominantly in coding regions. Perhaps methylation plays a role in regulating genes involved in developmental differences between honeybee queens and workers [11]. Finally, it appears that more than 60 genes are duplicated specifically in the honeybee, including two genes for components of the insulin pathway [2]. This pathway regulates growth in other animals [12] and could be the means through which queen bees become bigger than workers [13]. These and/or other duplicated genes may be involved in caste or sex determination and differences, and/or in social interactions.

Comparative analyses revealed some other peculiar genomic features including the fact that the honeybee genome evolves at a much slower rate than the strongly derived genomes of flies and mosquitoes. This is evidenced by sequence identity, intron conservation, and gene loss relative to an ancestor common to insects and vertebrates [2]. The slow evolution

of the honeybee genome may be general to hymenopteran insects, whose haplo-diploid sex-determination system might purge deleterious mutations that would be masked in diploid individuals [2]. Alternatively, it could be due to the long generation time of social insects — queens typically can live many years [14] — and/or to the low effective population size resulting from a single individual monopolizing reproduction in the hive.

Interestingly, these characteristics may also be responsible for another idiosyncrasy of the honeybee, the very high average recombination rate of 5.7 recombinations per chromosome [15]. Such a high recombination rate has only been reported in one other species, the ant Acromyrmex echinator [16]. High recombination rate might thus be a characteristic of social insects, again perhaps a result of their typically long generation time and small population size. Alternatively, high recombination rates might have been selected as a means to increase genetic diversity among offspring.

Another striking finding was that the honeybee genes controlling circadian rhythm and telomere length are more similar to vertebrates than to *D. melanogaster* or *A. gambiae* [2]. *Drosophila* is considered the paradigm of insect biology. This clearly needs to be changed given the increasing evidence that many of *Drosophila*'s features, such as their early-acting axis specification genes, are highly derived and not characteristic of insects.

The honeybee genome sequence and attendant analyses and experiments open many avenues for future research. In anticipation of a finished genome, the BeeSpace project has begun the dissection of environmental and hereditary influences on brain gene expression in the context of defensive behavior and foraging using thousands of microarrays (G. Robinson, personal communication). Beyond division of labor and cooperation within a colony, it will be exciting to understand the molecular basis of the evolution of within-colony conflicts and their resolution. Functional tests, through ectopic expression or repression of genes involved in both behavior and caste development, will be essential in elucidating how insect societies function.

At a more basic level, the mechanisms by which solitary species became social could be pinpointed by examining either facultatively social species or contrasting pairs of solitary and primitively social species. The independent evolution of social life in bees, ants, wasps and termites also provides a unique opportunity to determine whether the convergent morphological, physiological and behavioral adaptations that have occurred in these taxa are due to modification of the same developmental pathways and gene networks. The recent development of an EST library, microarray and other molecular genetic tools for the fire ant, *Solenopsis invicta* (our group's unpublished data), should pave the way for such sociogenomic comparative studies.

2.1 BIBLIOGRAPHY

1. Hölldobler, B., and Wilson, E.O. (1990). The Ants. Belknap Press.

2. The Honeybee Genome Sequencing Consortium (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 443, 931–949.

3. Osborne, K.A., Robichon, A., Burgess, E., Butland, S., Shaw, R.A., Coulthard, A., Pereira, H.S., Greenspan, R.J., and Sokolowski, M.B. (1997). Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. Science 277, 834-836

4. Keller, L, and Ross K.G. (1998). Selfish genes: a green beard in the red fire ant. Nature 394, 573–575.

5. Whitfield, C.W., Ben-Shahar, Y., Brillet, C., Leoncini, I., Crauser, D., LeConte, Y., Rodriguez-Zas, S., and Robinson, G.E. (2006). Genomic dissection of behavioral maturation in the honey bee. Proc. Natl. Acad. Sci. 103, 16068–16075.

6. Sinha, S., Ling, X., Whitfield, C.W., Zhai C., and Robinson G.E. (2006). Genome scan for cis-regulatory DNA motifs associated with social behavior in honey bees. Proc. Natl. Acad. Sci. 103, 16352–16357.

7. Hummon, A.B., Richmond, T.A., Verleyen, P., Baggerman G., Huybrechts, J., Ewing, M.A, Vierstraete, E., Rodriguez-Zas, S.L., Schoofs, L., Robinson, G.E., and Sweedler, J.V. (2006). From the genome to the proteome: uncovering peptides in the *Apis* brain. Science 314, 647–649.

8. Drapeau, M.D., Albert, S., Kucharski, R., Prusko, C., and Maleszka, R. (2006). Evolution of the Yellow/Major Royal Jelly Protein family and the emergence of social behavior in honey bees. Genome Res. 16, 1385-1394.

9. Robertson, H.M., and Wanner, K.W. (2006). The chemoreceptor superfamily in the honey bee *Apis mellifera*: Expansion of the odorant, but not gustatory, receptor family. Genome Res. 16, 1395-1403.

10. Evans, J.D., Aronstein, K., Chen, Y.P., Hetru, C., Imler, J.L., Jiang, H., Kanost, M., Thompson, G.J., Zou, Z., Hultmark, D. (2006). Immune pathways and defence mechanisms in honey bees *Apis mellifera*. Insect Mol. Biol. 15, 645–656.

11. Wang, Y., Jorda, M., Jones, P.L., Maleszka, R., Ling, X., Robertson, H.M., Mizzen, C.A., Peinado, M.A., and Robinson, G.E. (2006). Functional CpG methylation system in a social insect. Science 314, 645–647.

12. Wu, Q., and Brown, M.R. (2006). Signaling and function of insulin-like peptides in insects. Ann. Rev. Entomol. 51, 1-24

13. Wheeler D.E., Buck, N., and Evans, J.D. (2006). Expression of insulin pathway genes during the period of caste determination in the honey bee, *Apis mellifera*. Insect Mol. Biol. 15, 597–602.

14. Keller, L., and Genoud, M. (1997). Extraordinary lifespans in ants: a test of evolutionary theories of ageing. Nature 389, 958-960

15. Beye, M., Gattermeier, I., Hasselmann, M., Gempe, T., Schioett, M., Baines, J.F., Schlipalius, D., Mougel, F., Emore, C., Rueppell, O, et al. (2006). Exceptionally high levels of recombination across the honeybee genome. Genome Res. 16, 1339-1344

16. Sirvio, A., Gadau, J., Rueppel, O., Lamatsch, D., Boomsma, J.J.,Pamilo, P., and Page, R.E. Jr. (2006). High recombination frequency creates genotypic diversity in colonies of the leaf-cutting ant Acromyrmex echinatior. J. Evol. Biol. 19, 1475–1485.

Part II

DEVELOPMENT OF TOOLS FOR ANT GENOMICS

3

FOURMIDABLE: A DATABASE FOR ANT GENOMICS

by Yannick Wurm, Paolo Uva, Frédéric Ricci, John Wang, Stephanie Jemielity, Christian Iseli, Laurent Falquet and Laurent Keller

Before my arrival in Lausanne, the Fourmidable database had already been set up by Paolo Uva and Laurent Falquet after discussions with John Wang and Stephanie Jemielity. However, due to changes in technology and end-user requirements it became necessary to rewrite the sequence cleaning, assembly and annotation pipelines. Additional, many usability improvements were implemented. These changes were determined and implemented by myself with technical assistance from Frédéric Ricci. For one step of the sequence assembly pipeline, we used a clustering algorithm developed by Christian Iseli.

ABSTRACT

Background

Fourmidable is an infrastructure to curate and share the emerging genetic, molecular, and functional genomic data and protocols for ants.

Description

The Fourmidable assembly pipeline groups nucleotide sequences into clusters before independently assembling each cluster. Subsequently, assembled sequences are annotated via Interproscan and BLAST against general and insect-specific databases. Gene-specific information can be retrieved using gene identifiers, searching for similar sequences or browsing through inferred Gene Ontology annotations. The database will readily scale as ultra-high throughput sequence data and sequences from additional species become available.

Conclusions

Fourmidable currently houses EST data from two ant species and microarray gene expression data for one of these. Fourmidable is publicly available at http://fourmidable.unil.ch

This article was published in *BMC Genomics* (2009) 10:5
DOI: 10.1186/1471-2164-10-5

3.1 BACKGROUND

Ants are important model species for sociobiology and behavioral ecology [1] . Life in an ant colony is marked by cooperation, but it also entails conflicts. Both aspects have been studied extensively to understand the prerequisites for social behavior and to test the kin selection theory (e.g. [2,3]). New molecular and genomic techniques are making it possible to identify the genes underlying social behavior in ants and other social insects [4-11] as well as other fascinating aspects of social life including self-organization, life-history evolution, division of labor, and developmental plasticity [12-15]. The extraordinary complexity and vast information content generated by modern genomic techniques can be overwhelming. To take full advantage of such techniques requires appropriate bioinformatics tools.

To provide a central repository for the emerging ant genomic data, we developed Fourmidable, a web-accessible, user-friendly tool. Fourmidable currently provides detailed assembly and annotation of nucleic acid sequences from ants, a repository for ant microarray experiments and a platform to share ant-specific molecular biology protocols.

3.2 CONSTRUCTION AND CONTENT

Formidable contains publicly available sequence and gene expression data for ants and analyses of these data (summarized in Table 1).

Computation and Database Design

Fourmidable analyses are carried out via custom Perl scripts and publicly available software. Annotation and assembly information is stored in a MySQL database while sequences and BLAST [16] results are kept in indexed text files for rapid retrieval while limiting database size. Data are stored separately for different species [see Additional file 1]. Computationally intensive tasks are parallelized on the Vital-IT high-performance computing cluster [17, Additional file 2]. Fourmidable should thus readily handle large amounts of additional data. Data is accessible to web users via a PHP/Apache-based interface hosted by the Swiss Institute of Bioinformatics.

Nucleotide Sequence Data Preparation, Assembly, and Annotation

Fourmidable currently processes nucleotide sequences for the red fire ant *Solenopsis invicta* and the black garden ant *Lasius niger*. When available, raw.ab1 or.scf trace files are converted via Phred [18] to FASTA nucleotide and quality score files. Additional sequences for which trace files cannot be obtained are downloaded from Genbank; quality score files are generated for these sequences with an arbitrary Phred quality score of 25. All input sequences are cleaned using Lucy [19], DUST [Tatusov and Lipman, unpublished], RepeatMasker [20] and CrossMatch [21] to respectively remove low-quality regions and sequences, low-complexity regions, interspersed repeats, and sequences from bacteria, organelles or cloning vectors. Cleaned sequences are then compared via reciprocal BLAST [16], and subsequently similar sequences are grouped into clusters. Within each cluster, sequences are independently assembled via CAP3 [22]. This circumvents memory constraints that CAP3 would face if attempting a global assembly with large numbers of sequences. The output from clustering and CAP3 assemblies are contigs (each is the consensus of several assembled sequences) and singlets (sequences that did not assemble with others). All sequences are subsequently annotated as follows. All sequences are compared to the non-redundant protein database [23] via BLASTX as well as to several insect-specific databases via

TBLASTX, BLASTX and BLASTN. Gene Ontology (GO) annotation of the strongest of the top five BLASTX hits to the non-redundant protein database is carried over to ant sequences.

To determine possible peptide sequences, we compute all six possible translations of transcriptome sequences with the potential to encode sequences longer than 30 amino acids. We do not use an *ab initio* gene prediction program such as ESTScan [24] because the sensitivity of such programs is limited by the absence of solid training data for ants. Instead, all potential open reading frames are annotated via Interproscan [25]. Some Interproscan hits directly provide GO annotation of ant sequences, complementing the BLASTX-inferred GO annotation mentioned above.

The BLAST, Interproscan and GO annotations are updated every two months or when new sequence data is added to the assembly pipeline.

At several steps during this assembly and annotation pipeline, bioinformatics software was run with parameters that differed from default parameters [see Additional file 3], as determined by using reduced test datasets.

Gene Expression

Fourmidable is linked to the GEDAI gene expression database [Robin Liechti, unpublished]. Storage and simple analysis using Bioconductor packages [26] is possible for data from single-color and two-color spotted microarrays, as well as for Affymetrix and Illumina microarrays.

3.3 UTILITY AND DISCUSSION

As of October 2008, Fourmidable contains nucleotide sequence data for the fire ant *S. invicta* and the black garden ant *L. niger* as well as gene expression data for *S. invicta*. Currently accessible data are summarized in Table 1. Sequencing, gene expression profiling, and genotyping data are rapidly expanding and will be added as they become publicly available. Fourmidable's home page [27] centralizes links and search facilities to access Fourmidable's data and tools.

Sequence Search

There are several manners of accessing sequence information in Fourmidable. First, single sequences can be searched by species as well as by partial identifiers for input sequences or assembled contigs. Second, lists of identifiers can be used for searching. Third, user-supplied sequence data can be used for BLAST similarity searches against sequences in Fourmidable. Finally, users can navigate inferred Gene Ontology annotations for biological processes, cellular components, and molecular functions using the AmiGO browser [28]. The first two search manners result in tables as described below. BLAST searching and GO browsing produce lists of sequence identifiers that can be used as inputs for the first two search manners.

Sequence Information

Sequence searches result in tables with one line per sequence for easy access to sequence annotation (see Figure 1). In particular:

- Clicking on an identifier in the "Raw Sequence" column provides information on how that sequence was obtained and allows users to download the raw sequence. Tracefiles can be viewed with the Baylor College of Medicine Trace Viewer [29] or downloaded.

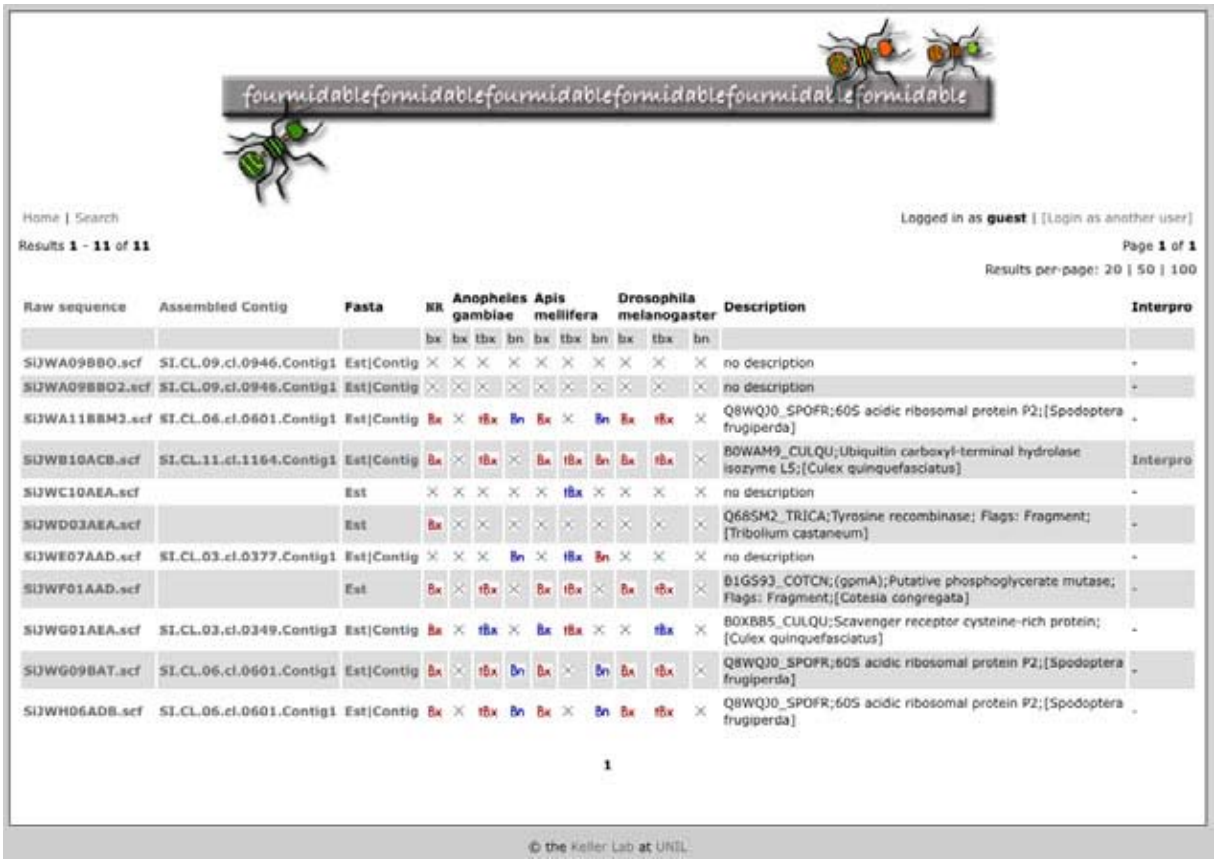


Figure 3.1: Table of sequence search results. For each result, the following are shown from left to right if applicable: sequence identifiers for raw and assembled sequence (these respectively link to the raw datafiles and assembly information); links to raw and assembled sequence in Fasta format; links to the results of BLAST against different databases (red buttons if E-value $< 10^{-5}$; blue buttons if 0.01 > E-value > 10^{-5}; bx, tbx and bn respectively indicate BLASTX, TBLASTX and BLASTN algorithms); a description as inferred from BLASTX against the non-redundant protein database; and a link to Interproscan annotation.

Solenopsis invicta:
- 28,006 input sequences including:
  - Tracefiles from 21,715 ESTs (some were multiply sequenced)
  - 1,496 additional ESTs and mRNA sequences from GenBank
- 12,859 putative transcripts:
  - 4,958 contigs
  - 7,263 singlets
- Sequence annotation:
  - 14,222 annotating GO terms on 2,818 putative transcripts
  - 599 Interproscan annotations
  - Blast comparisons against the non-redundant protein database, as well as proteomes and genomes of Apis mellifera, Anopheles gambiae and Drosophila melanogaster.
- Microarray data:
  - Two public experiments
  - 66 hybridizations

Lasius niger:
- 709 input sequences which are:
  - Tracefiles from 615 EST clones
- 403 putative transcripts:
  - 147 contigs
  - 256 singlets
- 21 Interproscan annotations

General:
- 8 molecular biology protocols

Table 3.1: Data Content in Fourmidable (October 2008)

- Cleaned FASTA-format sequence can be downloaded for individual singlets and contigs.
- If the sequence is part of an assembled contig, the "Assembled Contig" identifier links to a display of the consensus sequence and the relevant input sequences as well as their quality scores. Additionally, a multiple sequence alignment highlights nucleotide polymorphisms within the consensus sequence.
- BLAST results between the sequence of interest and sequences from the non-redundant protein database and several insect nucleotide and protein databases are summarized by blue and red buttons, indicating weak ($0.01 > \text{E-value} > 10^{-5}$) and stronger similarity ($\text{E-value} < 10^{-5}$) respectively. Clicking on a button displays the complete BLAST report.
- An additional link to Interproscan results and six-frame protein translations is displayed if Interproscan annotation is available.

Additional Features

A convenient repository is available for ant molecular biology protocols (commonly in.doc or.pdf formats). New or revised protocols can be added via an upload form. Fourmidable also supports upload of result files from microarray gene expression experiments. The GEDAI platform allows straightforward sharing of microarray results and performing simple microarray analyses (including preprocessing, direct and indirect two-sample comparisons, 2x2 factorial and gene set enrichment analyses). GEDAI also provides summaries of the expression levels of specific microarray probe identifiers across multiple microarray experiments. Finally, Fourmidable provides download links to individual files containing all raw or assembled sequences, as well as sequence annotation in text format [see also Additional file 4].

Past Applications

The sequence assembly and annotation information provided by Fourmidable has already proved useful in several published studies [14,30]. Most recently, Fourmidable's data helped J. Wang and colleagues to characterize genes that are differently expressed between workers from two alternative social forms of fire ants [6].

Outlook

Fourmidable was initially developed as a private database in Lausanne. Recently it has been updated and made publicly accessible because of increasing interest in ant molecular research. To further develop Fourmidable, several primary investigators in the USA have submitted grant applications. This should lead to improved integration of gene expression data with sequence annotation, as well as support for genetic mapping and linkage data. When large amounts of genomic sequence become available for ants, the current approach for assembly and annotation may become computationally unrealistic. An alternative may be to adapt existing genome assembly, annotation and browsing tools.

3.4 CONCLUSIONS

Fourmidable is a web-based database centralizing genomic resources for ants. As of October 2008, it contains raw sequence, assembled sequence, expression and annotation data for the fire ant *S. invicta* and the black garden ant *L. niger*, as well as ant-specific molecular biology protocols. Fourmidable will readily expand to accommodate additional data from these and additional species.

AVAILABILITY AND REQUIREMENTS

Fourmidable is publicly available [27]. It has been tested with Firefox 2 and 3, Safari 3 and Internet Explorer 7. The web interface is valid HTML 4.01 Transitional and CSS 2.1.

AUTHORS' CONTRIBUTIONS

LF, PU, YW, FR, SJ and JW designed the database. YW, PU, FR and CI developed the database. LK supported the work. YW drafted the manuscript. LK, LF, JW and SJ revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

We thank Marc Robinson-Rechavi and two anonymous reviewers for critical reading of the manuscript; Robin Liechti, Volker Flegel, Peter Sperisen, Lorenzo Cerruti and Vital-IT [17] at the Swiss Institute of Bioinformatics (http://www.isb-sib.ch) for advice and access to bioinformatics resources; Cedric Wurm for designing the icons used on Fourmidable's homepage.

This work was supported by the AR and J Leenards Foundation; the Swiss National Science Foundation; the Herbette Foundation; and the Rectorate of the University of Lausanne. LF is supported by the Swiss State Secretariat for Education and Research.

3.5 BIBLIOGRAPHY

1. Hölldobler B, Wilson EO: **The Ants**. 1990.
2. Sundstrom L, Chapuisat M, Keller L: **Conditional manipulation of sex ratios by ant workers: a test of kin selection theory.** *Science* 1996, **274**:993-995.
3. Bourke AFG, Franks NR: **Social Evolution in Ants**. 1995.
4. Robinson GE, Grozinger CM, Whitfield CW: **Sociogenomics: social life in molecular terms.** *Nat Rev Genet* 2005, **6**(4):257-270.
5. Robinson GE: **Integrative animal behavior and sociogenomics.** *TREE* 1999, **14**:202-205.
6. Wang J, Ross KG, Keller L: **Genome-wide expression patterns and the genetic architecture of a fundamental social trait.** *PLoS Genet* 2008, **4**:e1000127.
7. Keller L, Ross KG: **Selfish genes: a green beard in the red fire ant.** *Nature* 1998, **394**:573-575.
8. Whitfield CW, Cziko A, Robinson GE: **Gene expression profiles in the brain predict behavior in individual honey bees.** *Science* 2003, **302**:296-299.
9. Grozinger CM, Sharabash NM, Whitfield CW, Robinson GE: **Pheromone-mediated gene expression in the honey bee brain.** *Proc Nat Acad Sci USA* 2003, **100**(Suppl 2):14519-14525.
10. Nelson CM, Ihle KE, Fondrk MK, Page RE, Amdam GV: **The gene vitellogenin has multiple coordinating effects on social organization.** *PLoS Biol* 2007, **5**:e62.
11. Wurm Y, Wang J, Keller L: **Behavioral genomics: A, bee, C, G, T.** *Curr Biol* 2007, **17**:R51-R53.
12. Abouheif E, Wray GA: **Evolution of the gene network underlying wing polyphenism in ants.** *Science* 2002, **297**:249-252.
13. Jemielity S, Kimura M, Parker KM, Parker JD, Cao X, Aviv A, Keller L: **Short telomeres in short-lived males: what are the molecular and evolutionary causes?** *Aging Cell* 2007, **6**:225-233.
14. Graff J, Jemielity S, Parker JD, Parker KM, Keller L: **Differential gene expression between adult queens and workers in the ant Lasius niger.** *Mol Ecol* 2007, **16**:675-683.
15. Valles SM, Strong CA, Hunter WB, Dang PM, Pereira RM, Oi DH, Williams DF: **Expressed sequence tags from the red imported fire ant, Solenopsis invicta: annotation and utilization for discovery of viruses.** *J of Inv Pathol* 2008, **99**:74-81.
16. Altschul S, Madden T, Schaffer A, Zhang J, Zhang Z, Miller W, Lipman D: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
17. Swiss Institute of Bioinformatics: **Vital-IT center for high-performance computing** [http://www.vital-it.ch]
18. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8**:175-185.
19. Chou H, Holmes MH: **DNA sequence quality trimming and vector removal.** *Bioinformatics* 2001, **17**:1093-1104.
20. Smit A, Hubley R, Green P: **RepeatMasker Open-3.2.2** [http://www.repeatmasker.org]

21. **Cross_Match** [http://www.phrap.org/]

22. Huang X, Madan A: **Cap3: a DNA sequence assembly program.** *Genome Res* 1999, **9**:868-877.

23. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins.** *Nucleic Acids Res* 2005, **33**(Suppl 1):D501-504.

24. Iseli C, Jongeneel CV, Bucher P: **ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences.** *Proceedings of the International Conference on Intelligent Systems for Molecular Biology* 1999, 138-148.

25. Mulder NJ, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Buillard V, Cerutti L, Copley R, Courcelle E, Das U, Daugherty L, Dibley M, Finn R, Fleischmann W, Gough J, Haft D, Hulo N, Hunter S, Kahn D, Kanapin A, Kejariwal A, Labarga A, Langendijk-Genevaux PS, Lonsdale D, Lopez R, Letunic I, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Nikolskaya AN, Orchard S, Orengo C, Petryszak R, Selengut JD, Sigrist CJA, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C: **New developments in the Interpro database.** *Nucleic Acids Res* 2007, **35**(Suppl 1):D224-228.

26. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**:R80.

27. **Fourmidable** [http://fourmidable.unil.ch]

28. The GO Consortium: **The Gene Ontology (GO) project in 2006.** *Nucleic Acids Res* 2006, **34**(Suppl 1):D322-326.

29. Durbin KJ, Baylor College of Medicine Human Genome Sequencing Center: **Baylor College of Medicine Trace Viewer** [http://www.hgsc.bcm.tmc.edu/downloads/software/trace_viewer/]

30. Wang J, Jemielity S, Uva P, Wurm Y, Graff J, Keller L: **An annotated cDNA library and microarray for large-scale gene-expression studies in the ant Solenopsis invicta.** *Genome Biol* 2007, **8**:R9.

ADDITIONAL FILES

The additional files listed below can be downloaded from http://www.biomedcentral.com/1471-2164/10/5/additional/

File 1: **Notes on implementation.** We provide several details about decisions made relative to the implementation of the database, and the assembly pipeline.

On assembly:

In its current version, neither cluster nor contig identifiers generated during assembly are carried over from one assembly run to another. However, a tab-delimited text file is generated at the end of assembly run, showing correspondence between input sequences and contigs. Using this file, the history of which clones belong to which contig can be manually retraced.

On database implementation:

For each assembly "project" (meaning "species"), the MySQL database contains summary information referencing sequence, assembly, and annotations. Neither sequence nor annotation data are stored in the database, thus streamlining its size. Input data and cluster assembly information is stored in individual files; assembled sequence and blast results are respectively stored in indexed FASTA files and BLAST report files. All files are stored in project-specific directory hierarchies.

Adding data for a new species is straightforward: a new species identifier is created in the database, and a new directory hierarchy is created on disk.

File 2: **List of tasks parallelized on the Swiss Institute of Bioinformatics Vital-IT computing cluster.** Some tasks were parallelized for increased execution speed.

- RepeatMasker
- CrossMatch
- Cap3
- Interproscan
- Blast searches

File 3: **List of software parameters that differ from default**. For the assembly and annotation pipelines, default parameters were sometimes unsatisfactory. This table summarizes the parameters used when they differed from default.

Program	Parameter
SeqClean	-l 50
RepeatMasker	-lib arthropodRepeats.fasta *this "custom" repeats library is a download of all Arthropod and shared (ancestral) repeats from RepBase Update (23) in August 2008*
CrossMatch	-minscore 20 -masklevel 100
CAP3	-d 400 -o 21 -s 900
BLAST *(for clustering)*	-p blastn -q -3 -G 2 -E 4 -K 0 -v 1 -b 1000 -e 1e-4
BLAST *(for annotation)*	-v 200 -b 5 -e 0.01
Interproscan	-trlen 30 -goterms -iprlookup -appl blastprodom fprintscan hmmpfam hmmsmart profilescan

File 4: **List of data available in text format.** Some of the data in Fourmidable can be downloaded in text format.

- Raw, cleaned, and assembled sequences in FASTA format
- Assembly report
- BLAST results
- Gene Ontology annotation

4

AN ANNOTATED CDNA LIBRARY AND MICROARRAY FOR LARGE-SCALE GENE-EXPRESSION STUDIES IN THE ANT *SOLENOPSIS INVICTA*

by John Wang*, Stephanie Jemielity*, Paolo Uva, Yannick Wurm, Johannes Gräff and Laurent Keller

My first major contribution to this paper was to compare the electronically inferred Gene Ontology annotations of putative S. invicta *genes to the annotations of genes in the* Drosophila *genome (See Table 2 and the* Functional annotation *section). My second major contribution was to identify sequences that appear to be shared exclusively by the fire ant and the honey bee (See Table 3, Figure 2, the* Being a Hymenopteran *section and* Additional Data File 5*).*

ABSTRACT

Ants display a range of fascinating behaviors, a remarkable level of intra-species phenotypic plasticity and many other interesting characteristics. Here we present a new tool to study the molecular mechanisms underlying these traits: a tentatively annotated expressed sequence tag (EST) resource for the fire ant *Solenopsis invicta*. From a normalized cDNA library we obtained 21,715 ESTs, which represent 11,864 putatively different transcripts with very diverse molecular functions. All ESTs were used to construct a cDNA microarray.

This article was published in *Genome Biology* (2007) 8:1, R9
DOI: 10.1186/gb-2007-8-1-r9

4.1 BACKGROUND

Ants are important model species for sociobiology and behavioral ecology [1]. Life in an ant colony is marked by cooperation, but it also harbors conflicts. Both aspects have been studied extensively to understand the prerequisites for social behavior and to test the kin selection theory (reviewed in [2]). Other fascinating research areas in ants include self-organization, life-history evolution, as well as division of labor.

With the advent of new molecular and genomic techniques it is becoming possible to identify the genes underlying social behavior [3,4], as well as those involved in other interesting behaviors and traits. Unfortunately, in ants such studies have been seriously constrained by the lack of sequence data and other molecular tools. The majority of ant gene sequences have derived from two studies. A recent experiment examined differential gene expression in fire ants between winged virgin queens and wingless mated queens [5]. From this study 81 expressed sequence tags (ESTs) were submitted to GenBank. Another study, focusing on gene expression changes during the development of Camponotus festinatus workers, yielded 384 ESTs [6]. While informative, both of these studies were limited by the small number of genes examined. The goal of this project was, therefore, to create and sequence a much larger set of ant ESTs, namely for the ant Solenopsis invicta. Used in conjunction with DNA microarray technology [7,8], this sequence resource will allow us and other researchers to examine thousands of ant genes simultaneously.

S. invicta is one of the most extensively studied ant species. Also known as the red imported fire ant because of its accidental introduction to the United States from South America in the early 1900s and because of its painful, burning sting, this species has become a major agricultural and wildlife pest in the southern USA [9]. In attempts to control this species, its basic biology has been well elucidated [10,11]. Studies on S. invicta led the way in a number of research areas important for evolutionary biology: nest-mate conflicts over reproduction [12,13], sex-ratio conflicts [14,15], nepotism [16], chemical communication and warfare [17,18], and social evolution [19]. A particularly fascinating aspect of fire ant biology is that two distinct types of social organization exist in this species, and this is linked to a single gene, Gp-9 [20-22]. Colonies of the monogynous form are headed by a single reproductive queen with a specific Gp-9 genotype (BB), while colonies of the polygynous form contain up to several hundred reproductive queens that are all Gp-9 heterozygotes (Bb). The number of queens is regulated by workers, which will kill or tolerate additional queens based on their own and the queens' Gp-9 genotype [22]. This is one of a few cases where a complex social behavior is governed by a simple genetic mechanism.

We describe here a collection of 21,715 S. invicta ESTs generated from a normalized cDNA library. This library should encompass a maximum variety of genes, as it was derived from mRNA of all developmental stages of queens, males and workers from both colony types. Sequence assembly resulted in 11,864 putatively different genes. We have used a combination of blast analysis and protein pattern searches to obtain a preliminary Gene Ontology (GO) annotation for these genes. By comparison to the honey bee, we identified 23 potential Hymenoptera-specific genes. All ESTs were used to generate a high-density cDNA microarray, which will be a valuable resource for molecular, ecological and evolutionary studies in ants.

Total number of sequence reads	28,133
cDNA clones sequenced from 5' end	22,560
Extra reads due to re-sequencing	5,573
High-quality sequences after filtering*	21,715
Average EST size after trimming (bp)	522.4
Total number of assembled sequences	11,864
Number of contigs	4,319
True contigs (from >2 different clones)	3,057
Re-sequencing contigs†	1,262
Number of singletons	7,545
Number of putatively different fire ant sequences	<11,864
Average size of assembled sequences (bp)	600.5

*High quality sequences are those with greater than 200 bp after trimming of vector and primer sequences and with a phred value higher than 15. In addition, this set excludes artifactual sequences that were manually removed. †Contigs composed of replicate sequences of only one clone

Table 4.1: Fire ant EST and assembly statistics

4.2 RESULTS AND DISCUSSION

Generation and assembly of fire ant ESTs

To survey the fire ant gene repertoire, we generated ESTs from a normalized cDNA library derived from ants of all developmental stages and castes (workers, queens, and males) of both the monogynous and polygynous social forms. First, we sequenced the 5´ ends of 22,560 clones from the cDNA library. This yielded a total of 28,113 sequence reads, since about one-fourth of all clones were sequenced twice. From this set we then removed artifactual sequences and sequences smaller than 200 base pairs (bp; after vector and primer clipping), identifying 21,715 high-quality ESTs of 522 bp average length (Table 1).

To find redundant transcripts, the 21,715 ESTs were assembled into contiguous sequences (contigs, Table 1) using the Paracel Clustering Package. A total of 14,170 ESTs were assembled into 4,319 contigs, while the remaining 7,545 ESTs remained singleton sequences. In sum, there were 11,864 gene sets, hereafter referred to as assembled sequences, that putatively represent different transcripts. However, this number is expected to overestimate the true number of transcripts represented because some non-overlapping ESTs may represent the same gene and because assembly may have failed in case of alternative splicing, sequence polymorphism or sequencing errors. Assessed with a second independent method, the number of putatively different fire ant transcripts was indeed estimated at 'only' 9,770 (see below). The average length of all assembled sequences was 600 bp.

Since some of the cDNA clones were sequenced several times, 1,262 of the 4,319 contigs are due to re-sequencing, that is, composed of sequences of a single re-sequenced clone. The remaining 3,057 contigs are 'true contigs', that is, derived from at least two independent cDNA clones (Table 1).

Quality of the cDNA clones and sequences

To obtain a tentative estimate of the percentage of 5´ truncated transcripts, we compared the fire ant assembled sequences to a set of 3,951 proteins listed on the eukaryotic orthologous groups (KOG) database [23] that are highly conserved among *Drosophila melanogaster*, *Caenorhabditis elegans* and *Homo sapiens*. In total, 1,827 fire ant assembled sequences had a highly significant blastx hit ($E \leq$ 1e-20) to the *Drosophila* KOG proteins. Among these, 749 (41%) had regions of similarity that started within the 20 first amino-terminal amino acid residues of their *Drosophila* homologs with either an in-frame

methionine at the same position as the fruitfly start methionine (588) or upstream of the alignment start (161). This suggests that up to 41% of the assembled sequences might have an intact 5′ end, whereas the remaining 59% are probably 5′ truncated.

The number of 3′ truncated transcripts was harder to estimate because most cDNA clones (52.8%) were not sequenced all the way through to their 3′ end (that is, the 5′ sequence reads were shorter than most cDNA clones). Nevertheless, since 39.3% of all fire ant ESTs ended with a polyA sequence, up to 39.3% of our ESTs may have an intact 3′ end. This is, however, likely to be an overestimate, as not all polyA sequences are true polyA tails.

Consistent with the expectation that the fire ant cDNA clones were sequenced from the 5′ end, 92.2% of all assembled sequences with significant similarity to a gene in the non-redundant (nr) database were encoded on the plus strand. This estimate was obtained by counting how many times the open reading frames (ORFs) of the fire ant assembled sequences matched that of their best homologs in other organisms (see next section). However, a small percentage of the ant assembled sequences (7.8%) appeared to be encoded on the minus strand. This could be due to non-specific annealing of the SMART adaptors, to transcription of an adjacent gene pointing in the opposite orientation, or to the presence of antisense transcripts in our library.

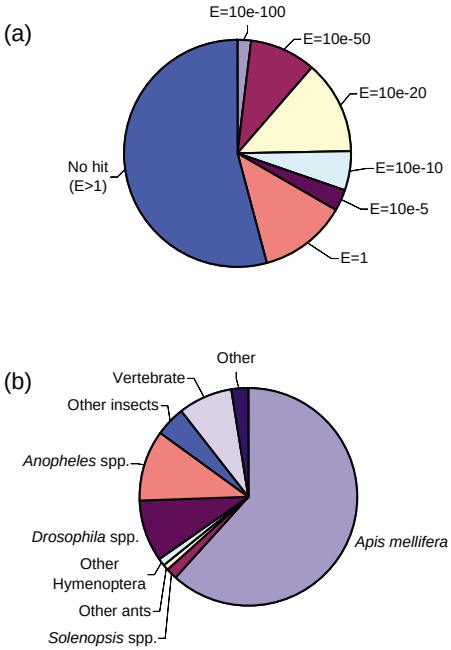


Figure 4.1: Sequence analysis by blastx searches. (a) Percentage of fire ant assembled sequences with and without blastx matches at various E-value cutoffs. (b) Quantitative overview of organisms providing the best-matching homologous protein sequences to fire ant assembled sequences (E $\leq$ 1e-5).

To assess overall sequence quality, we computed the number of unresolved bases, marked as N by the base-calling program phred, present in all ESTs and assembled transcripts. The majority of sequences (83.7% of assembled sequences and 81.3% of all ESTs) had no unresolved bases. Another 15.8% of assembled sequences and 17.5% of ESTs had between one and three unresolved bases. Finally, a small percentage of sequences (0.5% of assembled transcripts and 1.2% of ESTs) had more than four unresolved bases.

Comparative genomic analysis of fire ant cDNA data

We used the blastx algorithm to compare the 11,864 fire ant assembled sequences to the nr database. Of these, 2,936 (24.7%) and 3,964 (33.4%) assembled sequences matched known or predicted protein-coding genes at a cutoff expectation value (E) of 1e-20 and 1e-5, respectively (Figure 1a). By contrast, 6,431 (54.2%) had no similarity at all to genes in the nr database (E > 1). For many of these 6,431 clones, the lack of detectible similarity may be because the sequenced region does not encompass a long enough ORF to meet the blastx comparisons' cutoff of 1. This may result from 5′ truncation of cDNA clones (causing ESTs to consist mostly or entirely of 3′ untranslated region), from a long 5′ untranslated region, or from priming in intron regions of the pre-mRNAs.

Alternatively, transcripts may lack large ORFs because they are short or because they are noncoding RNAs (that is, transcripts other than rRNA or tRNA that do not code for proteins). Noncoding RNAs are now thought to make up a considerable portion of the polyadenylated transcripts found in libraries such as ours [24,25]. For instance, in humans 57% of all polyadenylated transcripts might be noncoding RNAs [26].

Figure 1b depicts the 'best hit' for the 3,964 fire ant assembled sequences displaying significant similarity to known or predicted protein-coding genes. The best hit was a honey bee gene 61.6% of the time. This was expected, as the honey bee is the most closely related species with a fully sequenced genome. Due to the paucity of non-honey bee hymenopteran sequences in GenBank, for only 106 (2.7%) assembled sequences was the best hit a known ant gene; and only 41 (1.0%) assembled sequences were most related to a gene from hymenopteran species other than ants or the honey bee. An additional 953 (24.0%) fire ant assembled sequences were most similar to genes from non-hymenopteran insect species. Of these, 359 and 417 had best matches to fruitfly and mosquito genes, respectively. Interestingly, a subset of 320 genes (8.1%) shared their closest similarity with vertebrates, which is an observation that has also been made for the honey bee [27]. Other assembled sequences were most similar to genes from Nematoda (11) or other Animalia (26). Several had best matches to bacteria (4) or protozoa (13), possibly because these sequences were derived from microbes that infect fire ants or that have a commensal relationship with them. Alternatively, these sequences could be due to microbial contaminations acquired during sample collection. Finally, 17 assembled sequences appeared to be derived from viruses, including the recently identified S. invicta SINV-1 and SINV-1A viruses [28,29].

Interestingly, for 1,341 fire ant assembled sequences the best hit was a non-hymenopteran gene (bacterial, viral and protozoan hits excluded). This could be due to extensive sequence divergence between ant-bee gene pairs or gene loss in the bee. We examined these two alternatives using the recently completed and annotated honey bee genome sequence [30]. Most fire ant genes with a non-hymenopteran best hit (72.880.5%; 1,080/1,341) had a significant blastx hit to an annotated honey bee gene (Additional data file 1). Using tblastx, blastn or Ensembl (v38 Apr 2006 [31]) honey bee gene predictions, an additional 69 fire ant genes showed evidence for a potential honey bee homolog (Additional data file 1). Thus, for these 1,149 assembled sequences, sequence divergence is the likely reason for a non-hymenopteran best hit. Such sequence divergence could be due to directional selection in the honey bee lineage. The remaining 192 (14.3%) assembled sequences do not display significant similarity to the honey bee genome (Additional data file 1). This could be because some ant sequences are too short to meet the significance threshold for similarity (1e-5), extreme sequence divergence, or putative gene loss in the honey bee lineage.

We also used the blastx analysis described as an alternative method to estimate the number of unique fire ant genes sequenced. A total of 3,366 fire ant assembled sequences matched 2,772 different honey bee proteins, suggesting that 82.4% (2,772/3,366) of the fire ant assembled sequences may be unique. Thus, the 11,864 fire ant assembled sequences may represent 9,770 different genes. Assuming that the fire ant and the honey bee have a similar total number of genes (that is, 13,448 to 20,998 predicted genes, Ensembl v38 April 2006 [31]), this would represent approximately 46.5% to 72.7% of the genes in the fire ant genome.

In addition to the above-mentioned blastx searches to identify putative protein-coding genes, we carried out two other genomic analyses. First, to identify potential noncoding RNAs among the fire ant assembled sequences, we compared all assembled sequences via blastn to known noncoding RNAs from the NONCODE database [32] and the miRBase microRNA collection [33]. Consistent with the view that noncoding RNAs are often poorly conserved across taxa [25], the vast majority of fire ant sequences had no significant hits in these databases (E > 1e-5). Only one fire ant transcript (SiJWG03CAD.scf) was highly similar (E = 3e-14) to a known human microRNA (miRBase ID: hsa-mir-594). Second, we identified 772 assembled sequences conserved between the fire ant and the honey bee that fulfilled

the following conditions: no resemblance to any known protein in the nr database (blastx, E > 1e-5), a good blastn hit against the honeybee genome (E ≤ 1e-5), and no significant blastn hit against other organisms (E > 1e-5). This list of genes (Additional data file 2) is likely to include transcripts with conserved untranslated region sequence motifs and some additional noncoding RNAs. However, it may also contain ant protein-coding genes that failed to have a blastx hit because they are truncated or because their honey bee homolog failed to be predicted during genome annotation.

Functional annotation

Provisional functional annotation of the fire ant assembled sequences was done by adopting the GO annotation of the best-matching homologues in the nr database. At a blastx E-value cutoff of 1e-5, 3,964 fire ant assembled sequences displayed matches to proteins in the nr database. Of these, 3,035 (76.6%) could be annotated into at least one of the three main GO categories (biological process, molecular function, or cellular component) and 1,617 (40.8%) were in all three. The distribution of the fire ant assembled sequences among the main subcategories is summarized in Table 2 and the full GO assignments are in Additional data file 3. The most frequently identified molecular functions were 'binding' and 'catalytic activity' and those for biological process were 'physiological process' and 'cellular process' (Table 2). In addition to the annotation through blastx searches, GO classifications were assigned to fire ant assembled sequences based on the Prosite protein domains they contain (Table 2, Additional data file 4). These two GO annotations were then contrasted with the GO annotation of the *D. melanogaster* genome: The relative counts of fire ant genes were significantly different (hypergeometric distribution: p < 1e-8) from the relative counts of *Drosophila* genes in up to 23 second-level GO categories (Table 2). This could indicate that these gene categories are over- or underrepresented in the fire ant genome relative to the *Drosophila* genome. Alternatively, these gene categories may simply be biased in cDNA libraries relative to genomes, for instance, because they contain mainly highly or mainly lowly expressed genes. GO groupings and subcategories can be further explored using the AmiGO feature [34] of the Fourmidable database. As the annotations are automated, all functional assignments are tentative and considered at the 'inferred from electronic annotation' (IEA) level of evidence (see [35]).

Being a Hymenopteran

The ants are classified within the order Hymenoptera, a group of insects including ants, bees and wasps. To identify Hymenoptera-specific genes, we looked for fire ant sequences that exhibited similarity only to genes from the honey bee or other Hymenoptera species. Using stringent criteria, we identified 148 fire ant sequences with strong similarity to the honey bee genome (tblastx, E < 1e-10) but no similarity to other known sequences (tblastx against non-hymenopteran sequences of the EMBL Nucleotide Sequence Database release 88; E > 1).

As the fire ant sequences are not necessarily full-length, the region of ant-bee homology, while apparently Hymenoptera-specific, may be part of a larger and phylogenetically conserved protein. To investigate this possibility, we examined the surrounding honey bee genomic sequence (±5,000 bp) of each candidate Hymenoptera-specific gene. Genes predicted by homology with other organisms were found near most of our putative ant-bee pairs. These regions of ant-bee homology may simply be fragments of known genes that diverged in ants and bees. However, for 23 ant-bee gene pairs (Table 3, Figure 2, Additional data file 5), the predicted neighboring genes are either specific to bees or are transcribed in the opposite direction. Unless the region of ant-bee homology is part of a conserved gene with a large intron (that is, >5,000 bp), these 23 ant-bee gene pairs are strong candidate Hymenoptera-specific genes.

	Solenopsis invicta EST library						D. melanogaster genome
	Blastx-determined GO			Prosite-determined GO			
Molecular function	4,301*		(100.0%)	486*		(100.0%)	14,778* (100.0%)
Antioxidant activity	20		(0.5%)	2		(0.4%)	39 (0.3%)
Binding	**1,765**	↑	**(41.0%)**	174		(35.8%)	4,319 (29.2%)
Catalytic activity	**1,456**	↑	**(33.9%)**	**201**	↑	**(41.4%)**	4,072 (27.6%)
Chaperone regulator activity	**5**	↑	**(0.1%)**	0		(0.0%)	1 (0.0%)
Enzyme regulator activity	91		(2.1%)	7		(1.4%)	382 (2.6%)
Molecular function unknown	**145**	↓	**(3.4%)**	**6**	↓	**(1.2%)**	1,852 (12.5%)
Motor activity	29		(0.7%)	1		(0.2%)	88 (0.6%)
Nutrient reservoir activity	**14**	↑	**(0.3%)**	0		(0.0%)	8 (0.1%)
Obsolete molecular function	0		(0.0%)	**9**	↑	**(1.9%)**	0 (0.0%)
Signal transducer activity	**153**	↓	**(3.6%)**	**4**	↓	**(0.8%)**	1,091 (7.4%)
Structural molecule activity	210		(4.9%)	59		(12.1%)	759 (5.1%)
Transcription regulator activity	**116**	↓	**(2.7%)**	4		(0.8%)	841 (5.7%)
Translation regulator activity	**62**	↑	**(1.4%)**	7		(1.4%)	92 (0.6%)
Transporter activity	235		(5.5%)	12		(2.5%)	1,014 (6.9%)
Triplet codon-amino acid adaptor activity	**0**	↓	**(0.0%)**	0		(0.0%)	220 (1.5%)
Cellular component	4,838*		(100.0%)	362*		(100.0%)	14,986* (100.0%)
Cell†	**1,868**	↑	**(38.6%)**	147		(40.6%)	5,225 (34.9%)
Cellular component unknown	**85**	↓	**(1.8%)**	**0**	↓	**(0.0%)**	1,920 (12.8%)
Envelope	107		(2.2%)	1		(0.3%)	290 (1.9%)
Extracellular matrix	14		(0.3%)	0		(0.0%)	46 (0.3%)
Extracellular matrix part	4		(0.1%)	0		(0.0%)	23 (0.2%)
Extracellular region	**73**	↓	**(1.5%)**	2		(0.6%)	416 (2.8%)
Extracellular region part	23		(0.5%)	0		(0.0%)	88 (0.6%)
Membrane-enclosed lumen	160		(3.3%)	3		(0.8%)	515 (3.4%)
Organelle	**1,360**	↑	**(28.1%)**	100		(27.6%)	3,007 (20.1%)
Organelle part	548		(11.3%)	22		(6.1%)	1,632 (10.9%)
Protein complex	575		(11.9%)	**87**	↑	**(24.0%)**	1,756 (11.7%)
Synapse	7		(0.1%)	0		(0.0%)	40 (0.3%)
Synapse part	3		(0.1%)	0		(0.0%)	27 (0.2%)
Virion†	**11**	↑	**(0.2%)**	0		(0.0%)	1 (0.0%)
Biological process	5,453*		(100.0%)	630*		(100.0%)	22,798* (100.0%)
Biological process unknown	**61**	↓	**(1.1%)**	**0**	↓	**(0.0%)**	888 (3.9%)
Cellular process	**2,242**	↑	**(41.1%)**	**297**	↑	**(47.1%)**	7,772 (34.1%)
Development	**121**	↓	**(2.2%)**	**0**	↓	**(0.0%)**	2,148 (9.4%)
Growth	17		(0.3%)	0		(0.0%)	102 (0.4%)
Interaction between organisms	6		(0.1%)	0		(0.0%)	92 (0.4%)
Physiological process	**2,328**	↑	**(42.7%)**	**315**	↑	**(50.0%)**	7,858 (34.5%)
Pigmentation	1		(0.0%)	0		(0.0%)	51 (0.2%)
Regulation of biological process	436		(8.0%)	11		(1.7%)	1,658 (7.3%)
Reproduction	**18**	↓	**(0.3%)**	**0**	↓	**(0.0%)**	826 (3.6%)
Response to stimulus	**207**	↓	**(3.8%)**	7		(1.1%)	1,402 (6.1%)
Viral life cycle	**16**	↑	**(0.3%)**	0		(0.0%)	1 (0.0%)

Listed are the numbers and percentages of assembled fire ant sequences and of D. melanogaster genes that match at least one of the second-level GO terms for molecular function, cellular component, or biological process. GO annotations for fire ant sequences were inferred electronically using two methods: blastx homology to GO-annotated proteins and Prosite protein domain scans. Statistically significant over- (↑) or underrepresentation (↓) of GO terms in fire ant relative to the Drosophila genome are indicated in bold ($p < 10^{-8}$, Bonferroni-corrected hypergeometric test). *This number represents the sum of the numbers of occurences of GO terms below this level. †The 'cell part' and 'virion part' GO categories were excluded from analyses because they were redundant with the 'cell' and 'virion' categories, respectively.

Table 4.2: Gene Ontology annotation

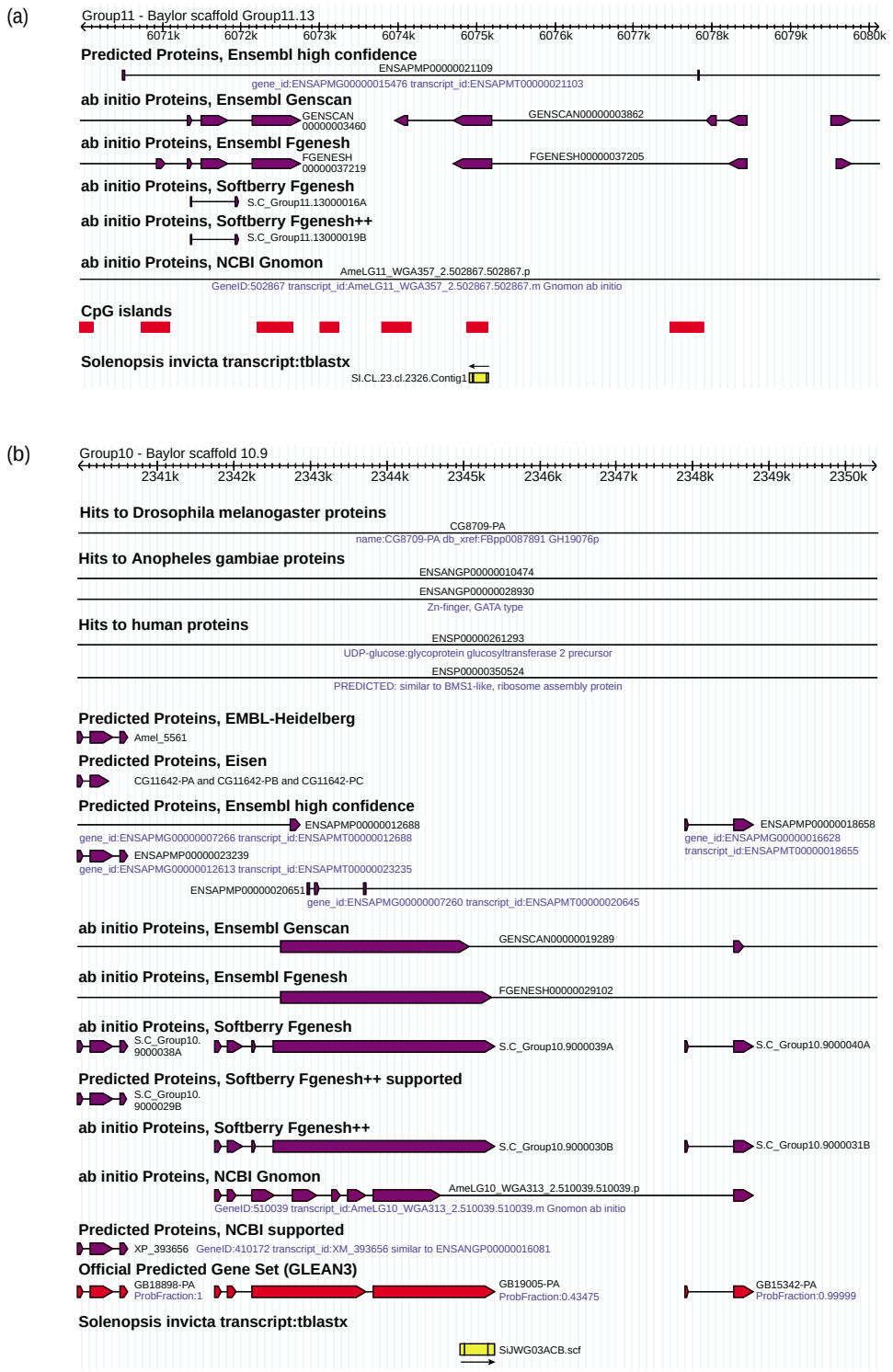


Figure 4.2: Examples of two candidate Hymenoptera-specific genes. (a) Fire ant sequence SI.CL.23.cl.2326.Contig1 matches an *ab intio* predicted honey bee gene that has no homology to any sequences in the public databases. The predicted gene was not included in the Honey Bee Official Gene Set. (b) Fire ant assembled sequence SiJWG03ACB.scf is the first EST evidence for the *ab initio* predicted honey bee gene GB19005-PA. Fire ant sequences are depicted as yellow boxes. Orientation (5' to 3') is indicated by an arrow. Predicted honey bee genes are depicted in purple; official Gene Set genes are shown in red. Images are based on output from Beebase (see Materials and Methods).

Further examination of these 23 candidate genes in hymenopteran species could prove interesting for understanding shared features. For instance, all Hymenoptera species have a haplodiploid sex determination system, with males developing from unfertilized haploid eggs and females from fertilized diploid eggs. Another feature found in many Hymenoptera is social behavior. Social behavior evolved independently in ants, bees and wasps [36,37] and, thus, it may be possible that a subset of the 23 ant-bee gene pairs was permissive for sociality to evolve or is important for social behavior.

Behavior genes

To identify candidate genes that might be involved in the complex behavior of ants we compared the fire ant assembled sequences to a set of 106 *Drosophila* genes that are directly implicated in behavior [27]. Of these behavior genes, 17 (16%) matched at least one fire ant assembled sequence (Table 4). This value is less than the 44% (47/106; chi-squared, $p < 5e\text{-}9$) identified by the honey bee brain cDNA library [27], possibly because the honey bee cDNA library was specifically derived from brain tissue. We also compared the fire ant assembled sequences to all 636 *Drosophila* genes that had the GO annotation 'behavior'. Of these, 81 (13%) were good hits for at least 1 fire ant assembled sequence (Additional data file 6). In addition, some genes involved in complex behaviors in ants and other Hymenoptera may be specific to this taxon and not homologous to known genes.

Viruses

In analyzing the cDNA library we noticed the presence of several viral transcripts. Seventeen fire ant assembled sequences were most similar to viral genes from RNA or DNA viruses (blastx, $E < 1e\text{-}5$; Table 5). Three sequences correspond to the recently identified SINV-1 virus, which possibly affects brood survival in *Solenopsis invicta* [28]. As the mutation rate in viruses can be high, we relaxed the E-value cutoff stringency to 1e-2, which yielded an additional nine putative viral genes. Based on different patterns of co-expression across several microarray experiments (unpublished data) the 26 putative viral genes could represent at least 5 different viruses.

To verify that these ESTs are from fire ant viruses and not from viruses infecting the insects fed to the ants, we tried to re-amplify all putative viral ESTs from fire ant cDNA derived from eggs, larvae and pupae. Out of 26 ESTs, 15 amplified when using egg and/or pupal cDNA as a template. Since eggs and pupae do not eat and either lack an intestine or have emptied their intestine, these 15 ESTs most likely stem from genuine fire ant viruses. Another five ESTs, including the three SINV-1 ESTs, amplified only in ant larvae. For these larvae-specific ESTs and the remaining six ESTs that amplified in none of the cDNA categories tested, additional tests would be needed to verify that they stem from fire ant viruses.

Further characterization of viruses in fire ants may be useful for two main reasons. First, as fire ants are an invasive pest species that causes considerable economic damage in the southern USA and other locations, viruses have been suggested as possible agents of fire ant control. Second, viruses can have dramatic effects on the behavior of their hosts. For instance, the Kakugo virus has been suggested to increase the aggressiveness of honey bee workers, as infected workers are much more likely to defend the nest against hornets than non-infected nestmates [38]. Another virus is most likely involved in superparasitism behavior in the parasitoid wasp Leptopilina boulardi [39]. It would be interesting to determine if the viruses identified by our EST project manipulate fire ant behavior to promote viral transmission or if they could be used for fire ant control.

Identifier (length)	*Solenopsis invicta* assembled sequence[1]				Blast statistics			*Apis mellifera* sequence				Confidence[7]
	Span	Frame	ORF[2] (bp)	f[3]	Exp[4]	Bit-score	E-value	Linkage group	Span	Strand	ORF[2] (bp)	Est[5]	Annotated gene[8]	
Sl.CL.8.cl.881.Contig1 (724 bp)	509-640	2	300		•	272	1.24E-18	6	27014427-2701558	+	429		*Ab initio* prediction	***
Sl.CL.8.cl.843.SiJWH04BDO2.scf (730 bp)	582-761	3	147		•	210	1.99E-12	NW_001254419.1[8]	44307-44486	-	147	•	Near NH homology. GB18184-PA on reverse strand	**
Sl.CL.19.cl.1938.Contig1 (835 bp)	21-323	3	372	T	•	212	1.43E-12	6	1145090-1145392	-	429		*Ab initio* prediction. Near GB12791-PA on reverse strand	***
Sl.CL.19.cl.1953.SiJWC11BBX.scf (613 bp)	81-215	3	555		•	166	5.08E-08	8	5253595-5253729	-	372	•	GB14543-PA. Near NH	*
	306-416	3	555		•	87	4.5E-15	8	5252894-5253094	-	306	•	homology on reverse strand	
Sl.CL.23.cl.2326.Contig1 (632 bp)	435-635	3	555		•	200	4.5E-15	8	5253189-5253299	-	318	•		
Sl.CL.23.cl.2328.Contig1 (859 bp)	413-577	2	219		•	291	1.33E-20	11	8022183-8022347	+	480	•	*Ab initio* prediction	***
Sl.CL.26.cl.2688.Contig1 (859 bp)	60-131	3	87		•	98	9.74E-15	9	10421877-10421948	-	549	•	*Ab initio* prediction. Near NH	**
	119-256	2	558		•	186	9.74E-15	9	10421751-10421888	-	549	•	homology on reverse strand	
Sl.CL.33.cl.3311.Contig1 (710 bp)	228-359	3	189		•	258	3.07E-17	14	8634060-8634191	-	132	•	Near *ab initio* prediction. Near NH homology on reverse strand	*
Sl.CL.33.cl.3384.Contig1 (469 bp)	229-327	1	264	T, S	•	160	3.11E-13	14	3770768-3770866	-	231		*Ab initio* prediction	***
	362-454	2	180	S	•	104	3.11E-13	14	3770649-3770741	-	186			
Sl.CL.35.cl.3595.Contig1 (415 bp)	123-398	3	342		•	301	5.97E-22	NW_001261806.1[8]	12471-12746	+	327		*Ab initio* prediction	***
SiJWA02BA22.scf (600 bp)	374-469	2	261		•	193	2.13E-15	5	9909503-9909598	+	627	•	Near GB15931-PA and NH	*
	533-604	2	261		•	98	2.13E-15	5	9909356-9909427	+	627	•	homology on reverse strand	
SiJWA03CAW.scf (666 bp)	49-144	1	96			120	2.1E-16	NW_001259848.1[8]	47860-47955	+	99	•	GB10007-PA on reverse strand	***
	136-297	1	117			182	2.1E-16	NW_001259848.1[8]	47704-47865	+	726	•		
SiJWA12ACK.scf (212 bp)	137-268	2	69		•	264	1.42E-19	3	5151467-5151598	+	162	•	Near *ab initio* prediction and NH	**
	63-143	3	72		•	69	1.42E-19	3	5151391-5151471	+	189	•	homology on reverse strand	
SiJWB12BCQ.tag5_B12_04.scf (754 bp)	121-369	1	354		•	254	1.1E-16	7	5620128-5620376	+	336		*Ab initio* prediction on reverse strand	***
SiJWC11BAT.scf (342 bp)	189-278	3	228		•	160	3.98E-17	14	8645843-8645932	+	162	•	Near *ab initio* prediction and homology	**
	282-368	3	228		•	123	6.41E-14	14	8645754-8645840	+	117	•		
SiJWE02BBO2.scf (865 bp)	714-863	3	129		•	243	1.26E-15	6	4850974-4851123	-	354	•	Near *ab initio* prediction on reverse strand	**
SiJWF07BCC.tag5_F07_11.scf (799 bp)	329-529	2	96		•	196	6.59E-11	3	6205208-6205408	-	108		Near NH homology. *Ab initio* prediction on reverse strand	**
SiJWG01BDU2.scf (759 bp)	21-227	3	102		•	354	1.23E-26	2	9618145-9618351	+	171	•	GB12576-PA and NH homology on reverse strand	*
SiJWG03ACB.scf (623 bp)	172-609	1	471		•	558	4.63E-47	10	2344965-2345402	+	1440	•	GB19005-PA	***
SiJWH02AAN.scf (469 bp)	100-294	1	102		•	341	1.32E-30	12	281374-281568	-	294	-		***
	28-105	1	69		•	104	1.32E-30	12	281564-281641	-	207			
SiJWH05BDPR5A08.scf (658 bp)	580-657	1	78		•	161	1.1E-15	10	2890267-2890344	+	159	•	Near *ab initio* prediction	**
SiJWH05BDV2.scf (517 bp)	204-353	3	198		•	237	4.87E-15	5	6704423-6704572	+	174		*Ab initio* prediction	**
SiJWH08AAT.scf (653 bp)	76-162	1	60		•	141	4.53E-20	5	1169177-1169263	+	84	•	Near *ab intrio* prediction and NH homology	*
	151-195	1	102		•	75	4.52E-13	5	1169261-1169305	+	69	•		
SiJWH08ADY.scf (563 bp)	236-496	2	327		•	312	1.32E-22	12	4477772-4478032	-	432	•	GB16574-PA	***

[1]*S. invicta* assembled sequences that show no significant similarity to any known non-hymenopteran sequence (E > 1), but high similarity to a region of the honey bee genome (E < e-10). [2]Length in base-pairs of the largest overlapping in-frame open reading frame. [3]In-frame Interproscan annotation of fire ant assembled sequence. T means 'transmembrane region', S means 'signal peptide'. [4]Gene is known (•) to be expressed in fire ant (unpublished microarray data). [5]In honey bee, EST evidence exists (•) within 5,000 bp of the aligned region. [6]This column shows the annotation of overlapping or nearby (within 5,000 bp) honey bee genes, as well as the nearby presence of genes from non-hymenopteran organisms. Numbers starting with GB are honeybee Official Gene Set numbers. '*Ab initio* prediction' indicates that Gnomon, Genscan, or another algorithm was used to predict a gene that was not retained for the bee genome Official Gene Set. 'NH homology' indicates the nearby presence of a gene from non-hymenopteran organisms. [7]Based on visual inspection we assigned a confidence level (the more asterisks the better) to each ant-bee putative gene pair (see Materials and methods). [8]*Apis mellifera* unanchored scaffolds such as NW_001254419.1 are regions that have not been mapped to a chromosome. [9]Multiple alignment frames for a *Solenopsis invicta* transcript indicate possible frameshifts during sequencing.

Table 4.3: Putative Hymenoptera-specific genes

Fire ant assembled sequence	Drosophila polypeptide ID	Gene name and behavior in Drosophila	E-value
SI.CL.10.cl.1087.Contig1	CG5670-PB	Na pump alpha subunit	1.0e-134
SI.CL.13.cl.1344.SiJWC08BDJ.scf	CG4443-PA	courtless (courtship behavior)	1.0e-73
SI.CL.13.cl.1344.Contig1	CG4443-PA	courtless (courtship behavior)	5.0e-73
SiJWE02ABO.scf	CG3263-PG	cAMP-dependent protein kinase R1 (olfactory learning)	4.0e-66
SiJWA12BCM.scf	CG2212-PA	swiss cheese	1.0e-65
SiJWC02AAC2.scf	CG3966-PA	neither inactivation nor afterpotential A	3.0e-55
SiJWB06ABV.scf	CG4379-PB	cAMP-dependent protein kinase 1 (locomotor rhythm, memory, olfactory learning and rhythmic behavior)	2.0e-42
SI.CL.3.cl.316.Contig1	CG8472-PB	calmodulin	2.0e-42
SI.CL.20.cl.2069.Contig1	CG2212-PB	swiss cheese	5.0e-42
SiJWH05AEA.scf	CG2048-PC	discs overgrown (altered behavioral response to cocaine)	4.0e-40
SiJWH06BAG.scf	CG8472-PB	calmodulin	4.0e-39
SI.CL.9.cl.956.Contig1	CG14724-PB	cytochrome c oxidase subunit Va	6.0e-38
SiJWA04BDS2.scf	CG3331-PA	ebony (locomotor rhythm)	7.0e-38
SiJWG01ADR.scf	CG7826-PC	minibrain (circadian rhythm and olfactory learning)	1.0e-24
SiJWD02ACW.scf	CG7758-PA	pumpless	1.0e-24
SI.CL.31.cl.3101.Contig1	CG1232-PB	temperature-induced paralytic E	3.0e-16
SiJWG06BCF2.scf	CG5670-PA	Na pump alpha subunit	8.0e-15
SiJWF02BDZ.scf	CG32688-PA	hyperkinetic (flight behavior)	1.0e-13
SiJWB11ABH.scf	CG10033-PG	foraging*	1.0e-11
SiJWB03ACL.scf	CG7100-PH	cadherin-N	2.0e-11
SiJWD03ACB.scf	CG10697-PA	aromatic-L-amino-acid decarboxylase (courtship behavior and learning and/or memory)	1.0e-07

*Although the best hit for SiJWB11ABH.scf is foraging, a type I cGMP-dependent protein kinase (PKG), when using blastx analysis with only the Drosophila predicted proteins, closer inspection using all the nr sequences suggests that it is actually a type II PKG.

Table 4.4: Fire ant assembled sequences putatively involved in behavior

Fire ant assembled sequence	Best virus hit ID	Hit description	E-value	Identity (%)
SI.CL.23.cl.2338.Contig1	Q5Y974	Structural polyprotein. [Solenopsis invicta virus 1]	0	98
SI.CL.23.cl.2338.Contig2	Q5Y974	Structural polyprotein. [Solenopsis invicta virus 1]	0	92
SI.CL.8.cl.873.Contig1	Q65353	ORF B. [Autographa californica nuclear polyhedrosis virus]	2.0e-76	52
SiJWG09BAM.scf	Q5Y975	Nonstructural polyprotein. [Solenopsis invicta virus 1]	2.0e-63	96
SiJWF01ADQ.scf	Q6AW71	(orf1)RNA-dependent RNA polymerase. [Bombyx mori Macula-like latent virus]	3.0e-51	93
SiJWB11ACS.scf	Q6AW71	(orf1)RNA-dependent RNA polymerase. [Bombyx mori Macula-like latent virus]	1.0e-44	90
SI.CL.29.cl.2930.Contig1	Q65353	ORF B. [Autographa californica nuclear polyhedrosis virus]	1.0e-43	55
SI.CL.28.cl.2823.Contig1	Q38QJ4	Polyprotein. [Kelp fly virus]	7.0e-34	28
SiJWC03CAP.scf	Q5ZNV0	Hypothetical protein. [Cotesia congregata bracovirus]	2.0e-22	51
SiJWA06BBH.scf	Q85431	RNA polymerase. [Rice stripe virus]	1.0e-21	35
SI.CL.37.cl.3723.Contig1	Q5S8C7	Non-structural polyprotein (Fragment). [Honey bee virus - Israel]	1.0e-18	40
SI.CL.41.cl.4135.Contig1	Q38QJ4	Polyprotein. [Kelp fly virus]	2.0e-15	34
SI.CL.19.cl.1909.Contig1	Q6AW70	(orf2)Coat protein. [Bombyx mori Macula-like latent virus]	2.0e-14	84
SI.CL.6.cl.610.Contig1	Q8QY61	Polyprotein. [Sacbrood virus]	2.0e-11	26
SI.CL.25.cl.2511.Contig1	O11437	(pv4)Non-capsid protein. [Urochloa hoja blanca virus]	6.0e-11	26
SI.CL.6.cl.610.Contig3	Q9QRA8	Polyprotein (Fragment). [Tomato ringspot virus]	2.0e-10	23
SI.CL.6.cl.610.Contig2	Q3YC01	Polyprotein (Fragment). [Stocky prune virus]	2.0e-06	29
SiJWA06CAM.scf	Q6QLR4	(RdRp)RNA-dependent RNA polymerase (Fragment). [Venturia canescens picorna-like virus]	3.0e-05	37
SiJWC05ADI.scf	Q5ZP67	Soluble protein. [Cotesia congregata bracovirus]	7.0e-05	38
SI.CL.40.cl.4005.Contig1	P03515	(N)Nucleocapsid protein (Nucleoprotein). [Punta toro phlebovirus]	4.0e-04	32
SiJWG01BBJ2.scf	Q9JGN8	(p1vc)P1. 339K. [Rice grassy stunt virus]	0.001	23
SiJWD07ACK.scf	Q8BDE0	Replicase polyprotein. [Acute bee paralysis virus]	0.002	25
SI.CL.10.cl.1089.Contig1	Q9YMJ7	Envelope protein. [Lymantria dispar multicapsid nuclear polyhedrosis virus]	0.003	23
SI.CL.16.cl.1675.Contig1	Q9YW13	(MSV079)Hypothetical protein MSV079. [Melanoplus sanguinipes entomopoxvirus]	0.004	42
SiJWH05ADG.scf	Q76LW4	Polyprotein. [Kakugo virus]	0.008	27
SiJWE11AAZ.scf	Q5ZNU9	Soluble protein. [Cotesia congregata bracovirus]	0.01	34

Table 4.5: Fire ant assembled sequences most similar to viral genes

Longevity

Ant queens and workers show up to ten-fold lifespan differences, although they develop from the same eggs and are thus genetically identical [1]. Lifespan differences must, therefore, stem from differences in gene expression, making ants a useful system to study aging and lifespan determination [40,41]. The average lifespan of fire ant queens is estimated at six to seven years [42], while workers are thought to have an average lifespan of ten to 70 weeks [1]. We have identified fire ant homologues (blastx, E < 1e-20) to several genes that are likely involved in determining the lifespan of invertebrate model organisms (reviewed in [43,44]): Cu-Zn superoxide dismutase (SI.CL.3.cl.379.Contig1), Mn superoxide dismutase (SI.CL.16.cl.1663.Contig1), catalase (SI.CL.40.cl.4085.Contig1), histone deacetylase Rpd3 (SiJWG06ABE.scf), Indy (SI.CL.40.cl.4047.Contig1) and the heatshock transcription factor HSF-1 (SiJWH04BCB2.scf). It will be exciting to test whether these homologues are expressed at different levels in the long-lived queens and the short-lived workers. In addition, comparing fire ant queens to fire ant workers using functional genomic approaches may help identify new candidate aging genes.

Highly expressed genes

In total, 67 contigs contained more than 10 ESTs (Additional data file 7). Consistent with the hypothesis that these are highly expressed genes, we found several homologs to ribosomal genes and other housekeeping genes in this subset. The largest contig (SI.CL.0.cl.071.Contig1) contained 48 clones. Based on blastx searches this gene encodes a small (74 amino acid residue) protein of unknown function. Interestingly, this gene is highly conserved across vertebrates, arthropods and fungi. For instance, the putative fire ant protein and its zebra fish homolog share 79% amino acid residues. While the majority of the 67 highly expressed transcripts had significant blastx matches to well-characterized proteins, 18 (26.9%) did not match any known sequence (E > 1e-5 for both blastx and blastn).

Fire ant microarray

To permit functional genomic analysis for the fire ant we produced a cDNA microarray using all 22,560 clones sequenced from the cDNA library. We successfully PCR-amplified 17,685 (78.4%) cDNAs (only one strong band, Additional data file 8), which putatively represent 10,12283 (85.3%) of the fire ant assembled sequences (Additional data file 9). To evaluate the percentage of cDNA spots derived from legitimate and sufficiently highly expressed transcripts, we examined the signal-to-background ratio of all spots in four test hybridizations (for details and additional analysis see Additional data files 10, 11 and 12). The two samples compared were derived from a mix of adults (workers, virgin queens, and males from both colony types in equal amounts) and a mix of brood (eggs, larvae and pupae of all castes in equal amounts). Of the spots derived from a single good PCR product, on average 82.8% (14,642/17,685) had an interpretable signal (that is, signal intensity greater than background plus two standard deviations), indicating that most cDNA clones are derived from legitimate transcripts.

Future prospects

The extraordinary complexity and diversity of morphology, behavior, and social organization in ants is far from being understood from a molecular genetics point of view. The present work, the largest collection of ESTs for an ant species, provides a valuable sequence, clone, and genomic resource for the ant research community. Using this resource it will be possible to identify genes important in caste

determination, behavioral genetics and plasticity, chemical communication, and population control. This microarray should also allow comparisons across related species. More broadly, as the genome sequence for the social honey bee, *Apis mellifera*, is available and that for the solitary wasp, Nasonia vitripennis, will soon arrive, comparisons and contrasts of both gene sequence and expression among the three species might shed light onto hymenopteran biology, behavior and social organization.

4.3 CONCLUSIONS

We have sequenced 22,560 ESTs from a normalized fire ant cDNA library and assembled them into 11,864 putatively unique transcripts. Using comparative genomic analyses and the GO vocabulary, we have functionally annotated the fire ant ESTs into a broad range of molecular functions and biological processes. Examination of the fire ant genes has led to the identification of 23 putative Hymenoptera-specific genes. Finally, we have developed a cDNA microarray that will be useful for large-scale gene expression profiling.

4.4 MATERIALS AND METHODS

Ants

Monogynous and polygynous fire ant colonies were collected in Georgia (USA) in 20032 and 20043 and transferred to the laboratory as previously described [45]. Colonies were maintained in climate-controlled rooms at 25°C and fed with crickets, mealworms, a mix of vegetables, and a mix of canned tuna fish, dog food and peanut butter. Samples were collected manually and immediately frozen in liquid nitrogen.

cDNA library

Using the Trizol reagent (Invitrogen, Carlsbad, CA, USA), total RNA was isolated from various samples of both monogynous and polygynous nests: eggs, small larvae, medium-sized larvae, sexual larvae, as well as pupae and adults of males, workers and queens (including both virgin and mated queens). We then pooled about 1 μg of each RNA sample to create a master sample with a maximum diversity of transcripts. This master sample was precipitated once with LiCl to eliminate contaminating DNA, quality checked on a 1% agarose gel and a Bioanalyzer 2100 chip (Agilent, Santa Clara, CA, USA) and sent in ethanol to Evrogen (Moscow, Russia) for cDNA library construction.

Evrogen constructed a normalized cDNA library using the SMART technology, which should enrich for full-length sequences. The plasmid used was pAL16. Based on PCR amplification of the inserts of 2,300 clones, the mean and median cDNA clone length was estimated at 940 bp and 850 bp, respectively. The shortest cDNA clone from this subset measured 180 bp, while the longest one measured about 3,300 bp. By comparison, the average *Drosophila* cDNA clone was 2 kb and the longest clone was 8.7 kb [46], suggesting that the fire ant cDNA library has many short clones that do not represent the entire transcriptional unit. Although the fire ant cDNA library is not directional, a 2 bp difference between the 3´ and 5´ SMART adaptors on all inserts permits sequencing cDNA clones specifically from the 5´ end.

Sequencing and sequence analysis

For 22,560 clones selected at random from the cDNA library, approximately 600 bp-sequence reads were obtained from the insert 5′ end. Of these clones, 5,573 were sequenced in duplicate (mostly both times from the 5′ end, with the exception of 77 clones that were sequenced from both the 3′ and the 5′ end). The primer used for the first approximately 8,000 sequences was SMART tag2 5′-AAGCAGTGGTATCAACGCAGAGTACG-3′ (which forms a 1 bp mismatch, in bold); the primer used for all other sequences was SMART tag2 fixed 5′-AAGCAGTGGTAACAACGCAGAGTACG-3′ (which matches perfectly). Sequencing was done by Synergene (Schlieren, Switzerland) on plasmid DNA extracted from overnight cultures. Base calling was performed with phred [47,48]. The Paracel Clustering Package (Paracel, Pasadena, CA, USA) was used to filter low-quality sequences (base calls with phred values <15 and EST length <200 bp), to remove vector and SMART adaptor sequence, as well as to mask polyA tails and other repetitive sequences. In addition, Paracel was used to identify and assemble redundant transcripts: ESTs that had an overlap of >50 bp were, when possible, automatically assembled into contiguous sequences (contigs). ESTs that did not meet this criterion were called singletons.

In order to find homologs of the fire ant assembled sequences in other organisms, all singletons and contigs were used to interrogate public sequence databases. Blast sequence alignments [49,50] were performed using the Blast Network Service provided by the Swiss Institute for Bioinformatics or on a desktop PC using standalone blast software. For both blastx and blastn searches the default settings were used. E-values are reported at 1e-5, except where indicated otherwise.

Gene Ontology annotation

We used the blastx algorithm to compare all 11,864 assembled sequences against the nr protein database. Using the best GO annotated SwissProt or TrEmbl hit with an E-value $\leq$ 1e-5, we annotated our transcripts at the IEA evidence level. Additionally, we scanned all assembled sequences for Prosite patterns with the stand-alone ps_scan perl program using the default cutoff level of 0 [51]. Transcripts having a Prosite pattern with a GO annotation were also annotated with the same GO terms at the IEA evidence level. In order to compare the fire ant GO annotations to those of *D. melanogaster*, we downloaded the *D. melanogaster* genome GO annotation from [52] on 19 September 2006. The WEGO web tool [53] was used to calculate the relative numbers of second-level GO categories within each first-level GO category (molecular function, biological process, cellular component) for both species. Using the hypergeometric test in R, we then tested which GO categories were significantly over- or underrepresented in the fire ant cDNA library relative to the *Drosophila* genome. Bonferroni correction was applied to the 80 tests carried out to correct for multiple comparisons.

Fourmidable database

A MySQL database with web interface was produced to house the fire ant EST and assembled sequence data (P Uva et al., manuscript in preparation). Users can view sequence trace files, perform blast searches against fire ant assembled sequences, download sequences, browse through blastx and GO annotations, and so on. The database is publicly accessible [54].

Identification of Hymenoptera-specific genes

All fire ant assembled sequences were compared against the nr protein database via blastx. The 6,948 transcripts that did not show strong similarity to the non-hymenopteran sequences of the nr database (blastx using BLOSUM45; E > 1) were subsequently aligned to the honey bee genome (build Amel 4.0). Of these, 216 ant transcripts had strong similarity to honey bee sequences (tblastx using BLOSUM45; E $\leq$ 1e-10). These 216 sequences were compared against all non-honey bee sequences of the EMBL Nucleotide Sequence Database (release 88, September 2006). We retained the 148 ant transcripts that showed strong similarity to honey bee build 4.0 (E $\leq$ 1e-10) and no or very weak similarity (E > 1) to known non-hymenopteran sequences (tblastx using BLOSUM45). When multiple tblastx alignment frames were possible, the positive strand frame with the strongest E-value was retained. The 10,000 bp honey bee genomic region surrounding each ant-bee sequence pair was then compared against the nr protein database via blastx. For 31 ant transcripts, the corresponding honey bee genomic region either did not show similarity to known genes, or only showed similarity to genes transcribed in the opposite direction. InterProScan was used to scan for protein signatures [55]. Additionally, the ant transcripts were aligned via tblastx against build 2.0 of the honey bee genome, which is currently the bee genome version with the most extensive annotation. With these results a GFF annotation file was generated and uploaded to BeeBase [56] for visual examination of all ant transcript-honey bee genome homolog pairs. Based on the existence and orientation of surrounding predicted genes we then determined a confidence level for each ant-bee pair. We assigned three stars when an ant transcript overlapped with a previously known bee gene (ab initio prediction or EST evidence); two stars if there was no known bee gene close by; one star if a gene from another organism appeared to hit within 5,000 bp of the ant-bee pair. In addition, 8 ant-bee pairs considered as false positives were eliminated, leaving us with 23 candidate Hymenoptera-specific genes. BeeBase was used to generate Additional data file 5 and a preliminary version of Figure 2, which was subsequently reformatted and modified to contain only relevant data: redundant text was removed, non-empty tracks were collapsed and empty tracks were deleted.

Microarray construction

Bacteria clones were inoculated into PCR plates containing 5 µl modified LB-ampicillin broth (0.2x LB and without NaCl) and grown overnight. Plasmid inserts were amplified by PCR after adding 95 µl of PCR mix. A single primer, SMART PCR primer 5´-AAGCAGTGGTAACAACGCAGAGT-3´, which matches both the 3´ and 5´ SMART adaptor of the inserts, was used. PCR mixes contained 0.4 µl 5 U/µl TAQ (Qiagen, Hilden, Germany), 10 µl 10× Qiagen buffer, 20 µl Q solution, 4 µl 25 mM $MgCl_2$, 1.5 µl 25 mM dNTPs, and 1 µl 100 µM SMART PCR primer. An initial 9 minute denaturation at 94°C was followed by 40 cycles of 30 s at 94°C, 30 s at 59°C, and 3 minutes at 72°C. The reaction ended with an additional incubation of 7 minutes at 72°C. PCR products (2 µl of each) were analyzed on a 1% agarose gel. Gel pictures were visually examined to classify all PCR products as follows: 'strong single band' (78.4%); 'no band' (3.9%); or 'weak or multiple bands' (17.5%). These data were used to create an Excel file (Additional data file 8), which will allow microarray users to exclude data from non-single-band spots. We preferred this solution to printing only single-band PCR products, as this would have involved an error-prone rearraying step.

PCR products were purified by a standard NaOAc/ethanol precipitation, resuspended in 30 µl water and transferred into duplicate 384-well plates using a Biomek FX liquid-handling robot (Beckman Coulter, Fullerton, CA, USA). Then PCR products were dried and resuspended in 20 µL 3× SSC, 1.5 M betaine. This spotting buffer improves spot homogeneity and signal-to-noise ratio [57]. We also resuspended 48 times 10 commercial exogenous controls (SpotReport Alien cDNA Array Validation

System, Stratagene, La Jolla, CA, USA) in 3× SSC, 1.5 M betaine, 1 set for each subgrid of the microarray. Microarrays were printed on aldehydesilane-coated slides (NexterionTM Slide AL, Schott Nexterion, Jena, Germany), using an OmniGrid 300 spotting robot (GeneMachines, San Carlos, CA, USA). Spot and printing quality were assessed visually under a dissecting microscope after printing. While a few slides had minor defects (for example, a few spots missing or damaged by dust particles), the majority of slides exhibited no defects. DNA was crosslinked to slides by baking at 80°C for 1 h. Afterwards, the slides were post-processed with $NaBH_4$ using the manufacturer's recommended protocol.

Clone tracking

To detect major mistakes (for example, inverted or rotated plates) made during sequencing, amplification and/or transfer into 384-well plates, we resampled and sequenced 534 PCR products from the 384-well plates. These samples were chosen so that they represented 2 to 4 samples of each 96-well plate. For all 96-well plates we also manually checked that PCR length patterns corresponded roughly to sequence length patterns. Using these 2 quality control methods, we identified 8 96-well plates that had been sequenced upside-down. After careful verification involving more sequencing, we corrected these mistakes by renaming the sequences correctly. At that point only 6 control sequences (1.1%) did not match the expected sequence, suggesting that these were sporadic contaminations.

Availability of sequence data, cDNA clones and microarrays

The ESTs described in this paper were submitted to the GenBank data library under accession numbers EE127747 to EE149461. The assembled sequences can be downloaded from the Fourmidable database [54]. The microarray data were submitted to Gene Expression Omnibus [58] with accession number GSE5995. Fire ant cDNA clones and cDNA microarrays can be obtained according to instructions on Fourmidable [54].

Additional data files

The following additional data are available with the online version of this paper. Additional data file 1 lists honey bee sequences similar to fire ant assembled sequences with a non-honey bee best hit. Additional data file 2 lists all fire ant assembled transcripts with a significant blastn hit to the honey bee genome and no other blastx or blastn hit. Additional data file 3 shows the GO annotations for all assembled transcripts based on blastx searches. Additional data file 4 shows the GO annotations for all assembled transcripts based on Prosite searches. Additional data file 5 shows the honey bee genome regions surrounding the candidate Hymenoptera-specific genes listed in Table 3. Additional data file 6 contains fire ant assembled sequences similar to *D. melanogaster* genes with the GO term 'behavior'. Additional data file 7 contains an annotated list of the most abundant transcripts. Additional data file 8 shows the PCR results for the cDNA clones deposited onto the microarray. Additional data file 9 shows which fire ant assembled sequences had at least one cDNA clone with a good (single-band) PCR product. Additional data file 10 gives details on the microarray analyses performed. Additional data file 11 lists the fire ant clones that are differentially expressed between adults and brood based on a 4-fold cutoff. Additional data file 12 lists the fire ant clones that are differentially expressed between adults and brood based on a t-test ($p < 0.001$).

ACKNOWLEDGEMENTS

We thank M Robinson-Rechavi, G Robinson and three anonymous reviewers for critical reading of the manuscript; K Ross for ant colonies; L Falquet, P Sperisen and Vital-IT at the Swiss Institute of Bioinformatics for advice and access to bioinformatics resources; C LaMendola for help with ant sampling, RNA collections and microarray hybridizations; C Bernasconi for running PCR gels; A Patrignani and R Schlapbach at the Functional Genomics Center Zürich (FGCZ) for access to their liquid-handling robot. Special thanks to Keith Harshman, Johann Weber, Sophie Wicker, Manuel Bueno, and Jérôme Thomas at the Lausanne DNA Array Facility (DAFL) for microarray fabrication, advice and access to software. This research is supported by the AR and J Leenards Foundation (Lausanne), the Swiss National Science Foundation, the Rub Foundation, the Agassiz Foundation, the Herbette Foundation, the Chuard-Schmid Foundation and a grant from the rectorate of the University of Lausanne.

4.5 BIBLIOGRAPHY

1. Hölldobler B, Wilson EO: The Ants. Berlin: Springer-Verlag; 1990.
2. Bourke AFG, Franks NR: Social Evolution in Ants. Princeton: Princeton University Press; 1995.
3. Robinson GE: Integrative animal behaviour and sociogenomics. Trends Ecol Evol 1999, 14:202-205.
4. Robinson GE, Grozinger CM, Whitfield CW: Sociogenomics: social life in molecular terms. Nat Rev Genet 2005, 6:257-270.
5. Tian HS, Vinson SB, Coates CJ: Differential gene expression between alate and dealate queens in the red imported fire ant, *Solenopsis invicta Buren* (Hymenoptera : Formicidae). Insect Biochem Mol Biol 2004, 34:937-949.
6. Goodisman MAD, Isoe J, Wheeler DE, Wells MA: Evolution of insect metamorphosis: A microarray-based study of larval and adult gene expression in the ant Camponotus festinatus. Evolution 2005, 59:858-870.
7. De Risi JL, Iyer VR, Brown PO: Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 1997, 278:680-686.
8. Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene-expression patterns with a complementary-DNA microarray. Science 1995, 270:467-470.
9. Williams DF, Oi DH, Porter SD, Pereira RM, Briano JA: Biological control of imported fire ants (Hymenoptera: Formicidae). Am Entomol 2003, 49:150-163.
10. Taber SW: Fire Ants. College Station: Texas A&M University Press; 2000.
11. Tschinkel WR: The Fire Ants. Cambridge: Harvard University Press; 2006.
12. Vargo EL: Mutual pheromonal inhibition among queens in polygyne colonies of the fire ant *Solenopsis invicta*. Behav Ecol Sociobiol 1992, 31:205-210.
13. Bernasconi G, Krieger MJB, Keller L: Unequal partitioning of reproduction and investment between cooperating queens in the fire ant, *Solenopsis invicta*, as revealed by microsatellites. Proc R Soc B 1997, 264:1331-1336.

14. Vargo EL: Sex investment ratios in monogyne and polygyne populations of the fire ant *Solenopsis invicta*. J Evol Biol 1996, 9:783-802.

15. Passera L, Aron S, Vargo EL, Keller L: Queen control of sex ratio in fire ants. Science 2001, 293:1308-1310.

16. De Heer CJ, Ross KG: Lack of detectable nepotism in multiple-queen colonies of the fire ant *Solenopsis invicta* (Hymenoptera: Formicidae). Behav Ecol Sociobiol 1997, 40:27-33.

17. Fletcher DJC, Blum MS: Pheromonal control of dealation and oogenesis in virgin queen fire ants. Science 1981, 212:73-75.

18. Klobuchar EA, Deslippe RJ: A queen pheromone induces workers to kill sexual larvae in colonies of the red imported fire ant (*Solenopsis invicta*). Naturwissenschaften 2002, 89:302-304.

19. Ross KG, Vargo EL, Keller L: Social evolution in a new environment: the case of introduced fire ants. Proc Natl Acad Sci USA 1996, 93:3021-3025.

20. Ross KG, Keller L: Genetic control of social organization in an ant. Proc Natl Acad Sci USA 1998, 95:14232-14237.

21. Krieger MJ: To b or not to b: a pheromone-binding protein regulates colony social organization in fire ants. Bioessays 2005, 27:91-99.

22. Keller L, Ross KG: Selfish genes: a green beard in the red fire ant. Nature 1998, 394:573-575.

23. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, et al.: The COG database: an updated version includes eukaryotes. BMC Bioinformatics 2003, 4:41.

24. Claverie JM: Fewer genes, more noncoding RNA. Science 2005, 309:1529-1530.

25. Mattick JS: The functional genomics of noncoding RNA. Science 2005, 309:1527-1528.

26. Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G, et al.: Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science 2005, 308:1149-1154.

27. Whitfield CW, Band MR, Bonaldo MF, Kumar CG, Liu L, Pardinas JR, Robertson HM, Soares MB, Robinson GE: Annotated expressed sequence tags and cDNA microarrays for studies of brain and behavior in the honey bee. Genome Res 2002, 12:555-566.

28. Valles SM, Strong CA, Dang PM, Hunter WB, Pereira RM, Oi DH, Shapiro AM, Williams DF: A picorna-like virus from the red imported fire ant, *Solenopsis invicta*: initial discovery, genome sequence, and characterization. Virology 2004, 328:151-157.

29. Valles SM, Strong CA: *Solenopsis invicta virus-1A* (SINV-1A): distinct species or genotype of SINV-1? J Invertebr Pathol 2005, 88:232-237.

30. Honeybee Genome Sequencing Consortium: Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature 2006, 443:931-949.

31. Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, et al.: Ensembl 2006. Nucleic Acids Res 2006, 34:D556-561.

32. Liu C, Bai B, Skogerbo G, Cai L, Deng W, Zhang Y, Bu D, Zhao Y, Chen R: NONCODE: an integrated knowledge database of non-coding RNAs. Nucleic Acids Res 2005, 1:D112-D115.

33. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res 2006, 34:D140-D144.

34. Gene Ontology Consortium: The Gene Ontology (GO) project in 2006. Nucleic Acids Res 2006, 34:D322-D326.

35. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000, 25:25-29.

36. Cameron SA, Mardulyn P: Multiple molecular data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera : Apinae). Syst Biol 2001, 50:194-214.

37. Carpenter JM: Phylogenetic relationships and the origin of social behaviour in the Vespidae. In The Social Biology of Wasps. Edited by Ross KC, Matthews RW. Ithaca, NY: Cornell University Press; 1991:7-32.

38. Fujiyuki T, Takeuchi H, Ono M, Ohka S, Sasaki T, Nomoto A, Kubo T: Novel insect picorna-like virus identified in the brains of aggressive worker honeybees. J Virol 2004, 78:1093-1100.

39. Varaldi J, Fouillet P, Ravallec M, Lopez-Ferber M, Bouletreau M, Fleury F: Infectious behavior in a parasitoid. Science 2003, 302:1930.

40. Parker JD, Parker KM, Sohal BH, Sohal RS, Keller L: Decreased expression of Cu-Zn superoxide dismutase 1 in ants with extreme lifespan. Proc Natl Acad Sci USA 2004, 101:3486-3489.

41. Jemielity S, Chapuisat M, Parker JD, Keller L: Long live the queen: studying aging in social insects. AGE 2005, 27:241-248.

42. Tschinkel WR: Fire ant queen longevity and age - estimation by sperm depletion. Ann Entomol Soc Am 1987, 80:263-266.

43. Kenyon C: The plasticity of aging: Insights from long-lived mutants. Cell 2005, 120:449-460.

44. Tatar M, Bartke A, Antebi A: The endocrine regulation of aging by insulin-like signals. Science 2003, 299:1346-1351.

45. Jouvenaz DP, Allen GE, Banks WA, Wojcik DP: Survey for pathogens of fire ants, *Solenopsis* spp., (Hymenoptera-Formicidae) in the southeastern United States. Florida Entomol 1977, 60:275-279.

46. Stapleton M, Carlson J, Brokstein P, Yu C, Champe M, George R, Guarin H, Kronmiller B, Pacleb J, Park S, et al.: A *Drosophila* full-length cDNA resource. Genome Biol 2002, 3:RESEARCH0080.

47. Ewing B, Hillier L, Wendl MC, Green P: Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 1998, 8:175-185.

48. Ewing B, Green P: Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res 1998, 8:186-194.

49. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25:3389-3402.

50. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic Local Alignment Search Tool. J Mol Biol 1990, 215:403-410.

51. Gattiker A, Gasteiger E, Bairoch A: ScanProsite: a reference implementation of a PROSITE scanning tool. Appl Bioinformatics 2002, 1:107-108.

52. Gene Ontology [http://www.geneontology.org/]

53. Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L, et al.: WEGO: a web tool for plotting GO annotations. Nucleic Acids Res 2006, 34:W293-W297.

54. Fourmidable Ant Sequence Database [http://fourmidable.unil.ch/]

55. Zdobnov EM, Apweiler R: InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics 2001, 17:847-848.

56. BeeBase [http://racerx00.tamu.edu/cgi-bin/gbrowse/bee_genome2_chromo]

57. Diehl F, Grahlmann S, Beier M, Hoheisel JD: Manufacturing DNA microarrays of high spot homogeneity and reduced background signal. Nucleic Acids Res 2001, 29:E38.

58. Gene Expression Omnibus [http://www.ncbi.nlm.nih.gov/geo/]

ADDITIONAL DATA FILES

We have included additional data files with less than 500 lines within this document. All additional data files described below can be obtained at the following addresses:

- http://fourmidable012007.vital-it.ch/doc/sup_mat/index.html
- http://genomebiology.com/2007/8/1/R9/additional/

Additional data file 1 Honey bee sequences similar to fire ant assembled sequences with a non-honey bee best hit.

Additional data file 2 All fire ant assembled transcripts with a significant blastn hit to the honey bee genome and no other blastx or blastn hit

Additional data file 3 GO annotations for all assembled transcripts based on blastx searches

Additional data file 4 GO annotations for all assembled transcripts based on Prosite searches

Additional data file 5 Honey bee genome regions surrounding the candidate Hymenoptera-specific genes listed in Table 3

Additional Data File 5: Putative Hymenoptera-specific Genes
The 10,000 bp region around the ant-bee alignment of every putative Hymenoptera-specific gene is depicted relative to the annotated honey bee genome.

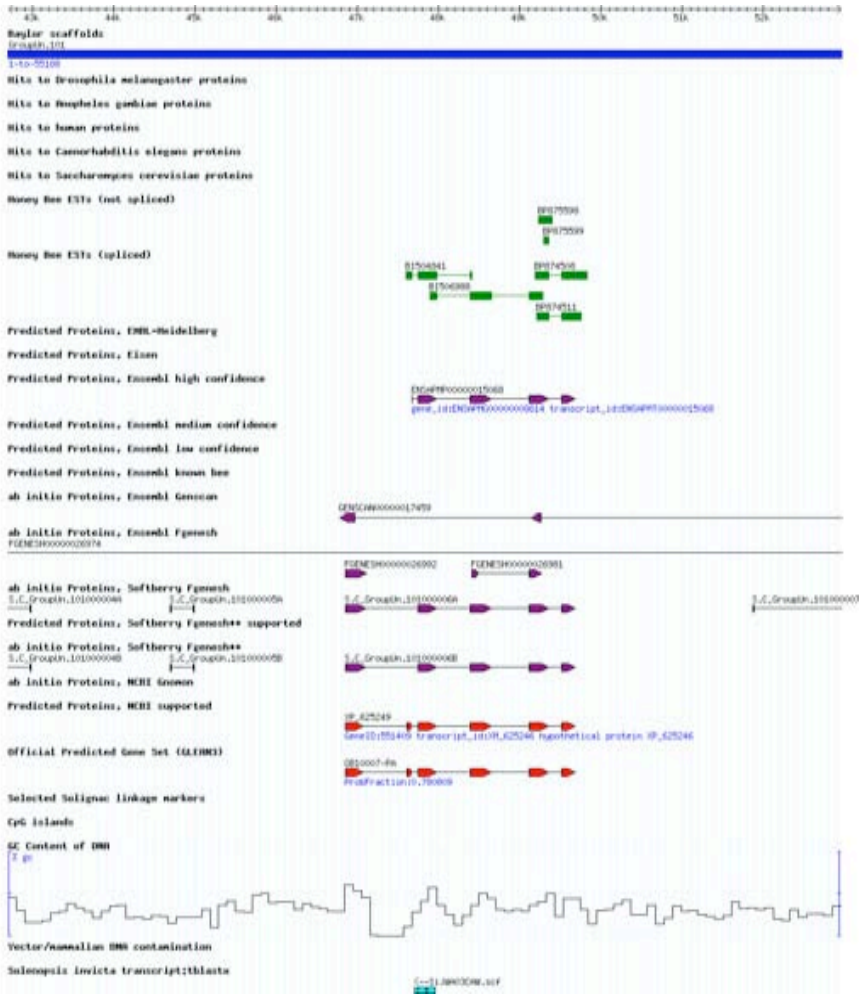
The honey bee genome assembly 2.0 chromosome coordinates are at the top of each image. Below, the alignment of proteins from *A. gambiae, C. elegans, D. melanogaster, H. sapiens* and *S. cerevisiae* are shown. Additionally, honey bee proteins predicted by several different groups (Eisen, EMBL, Ensembl, NCBI) as well as with multiple *ab initio* prediction methods are displayed. Finally, genes from the honey bee Official Gene Set (GLEAN3) are in red. The alignment of fire ant transcripts are in light blue; 5' to 3' direction is indicated by an arrow.

Coordinates may differ from those in Table 3 because Table 3 was generated using assembly 4 of the honey bee genome.

Index of subfigures:

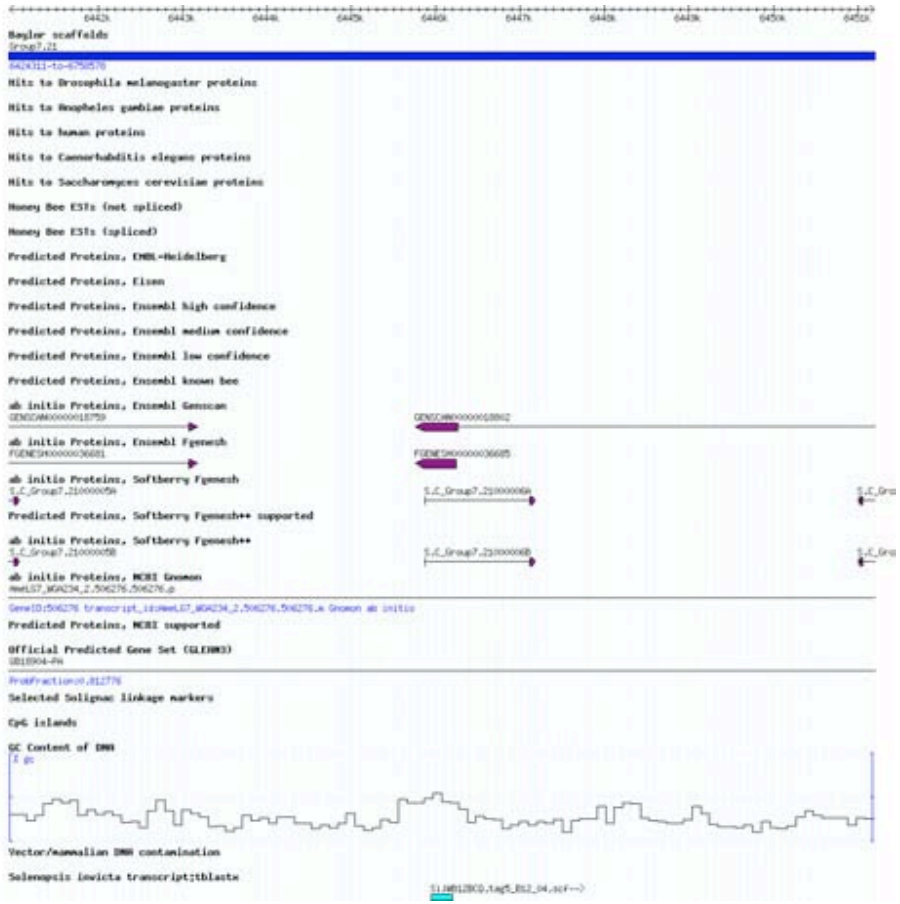
a.SiJWA02BAZ2.scf

b.SiJWA03CAW.scf



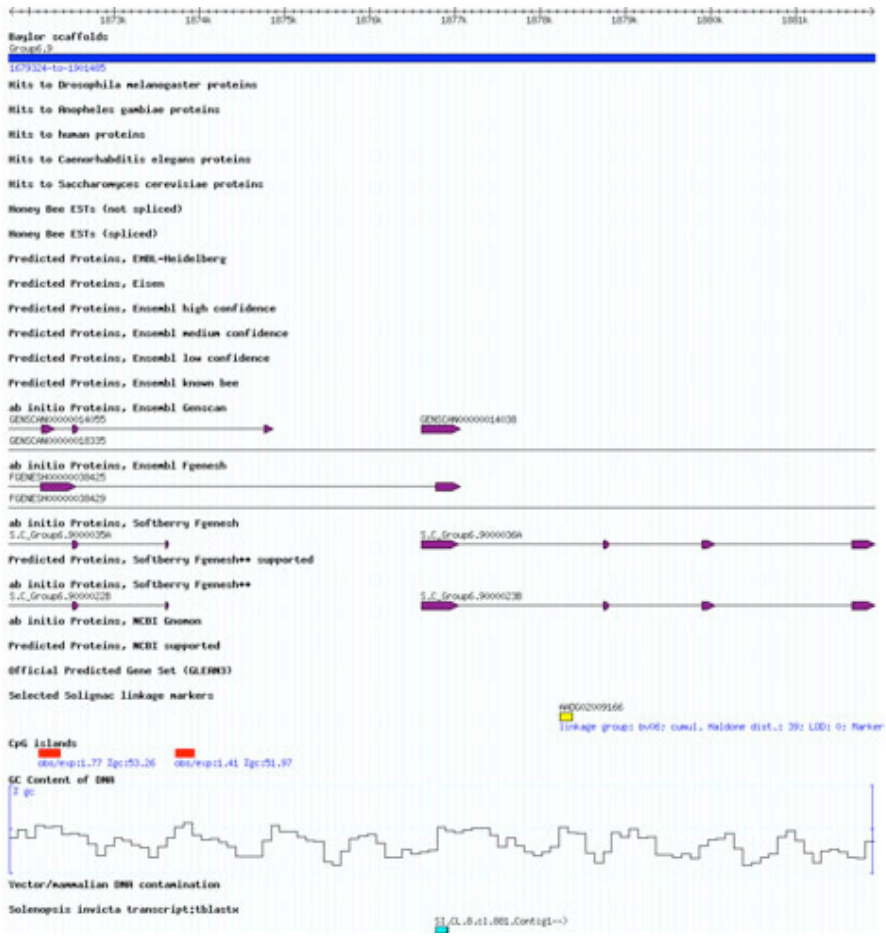
c.SiJWA12ACK.scf

d.SiJWB12BCQ.tag5_B12_04.scf



e.SiJWC11BAT.scf

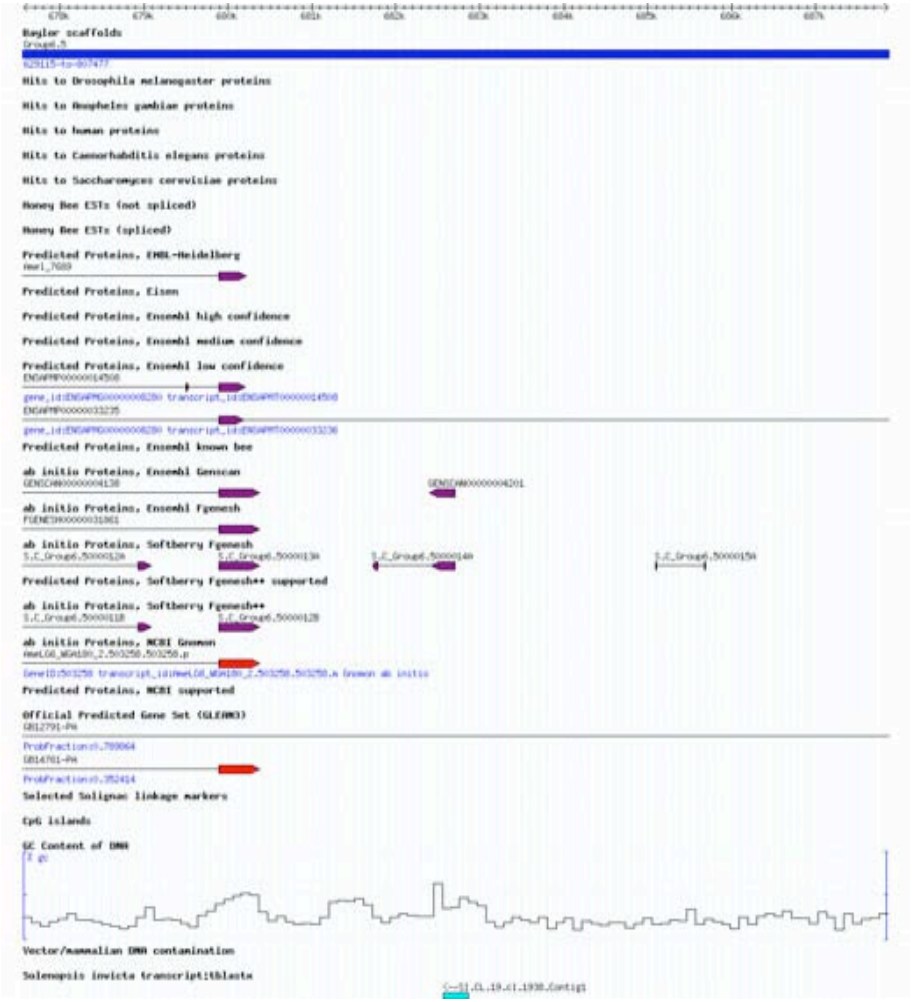
f. SI.CL.8.cl.881.Contig1



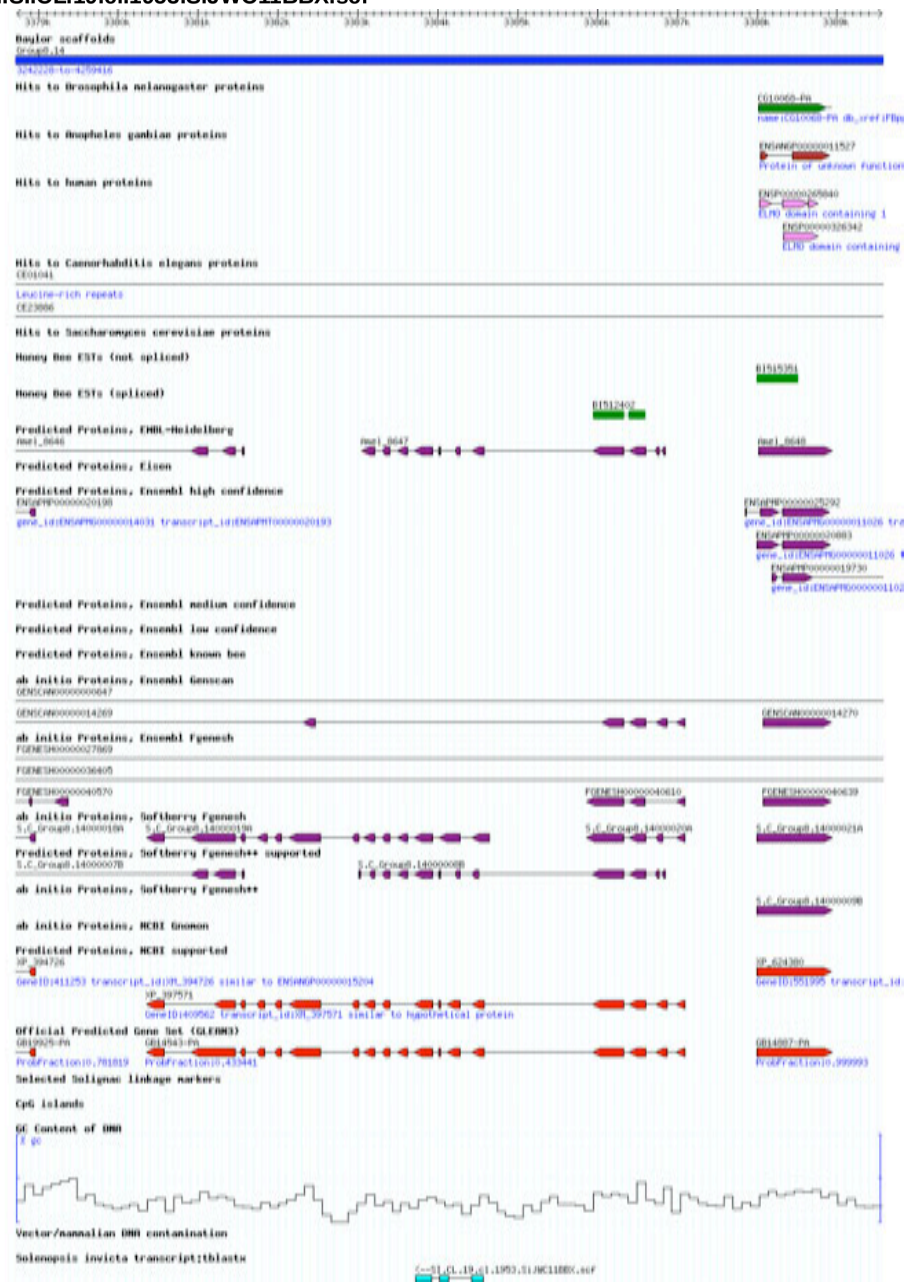
g.SI.CL.8.cl.843.SiJWH04BDO2.scf

* SI.CL.8.cl.843.SiJWH04BDO2.scf aligns close to one extreme end of the unanchored genomic scaffold GroupUn.742, which is why there is only little annotation data to the right side of SI.CL.8.cl.843.SiJWH04BDO2.scf.

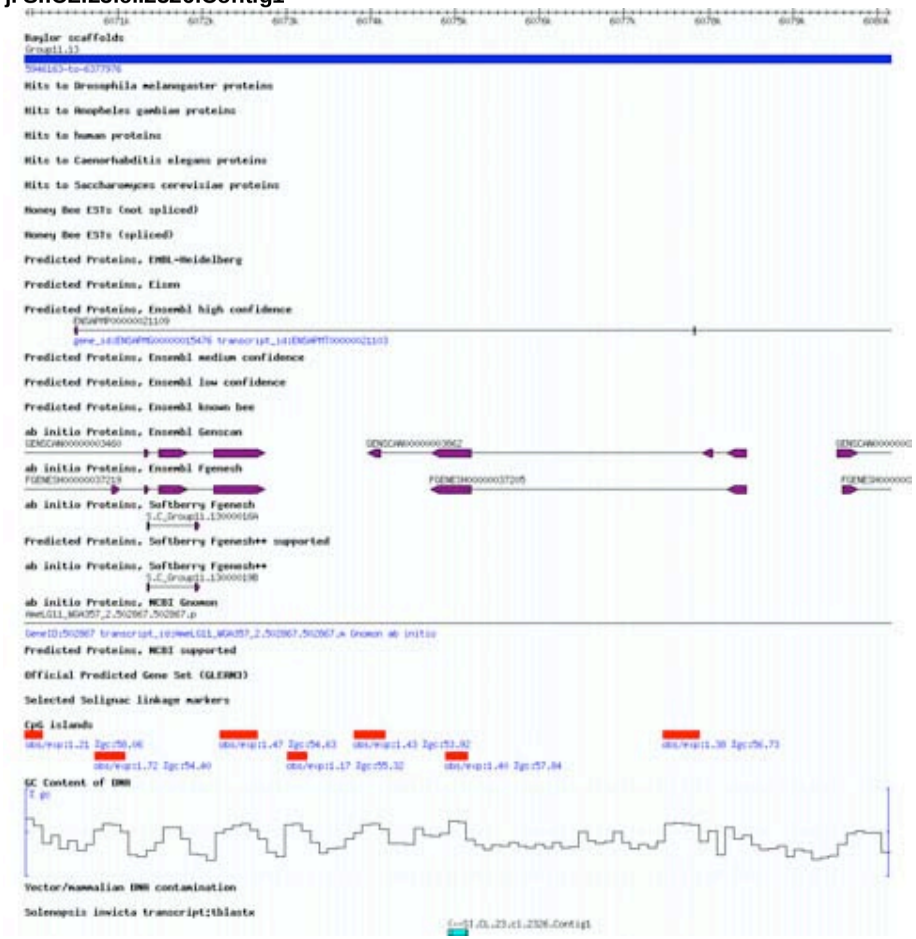
h. SI.CL.19.cl.1938.Contig1



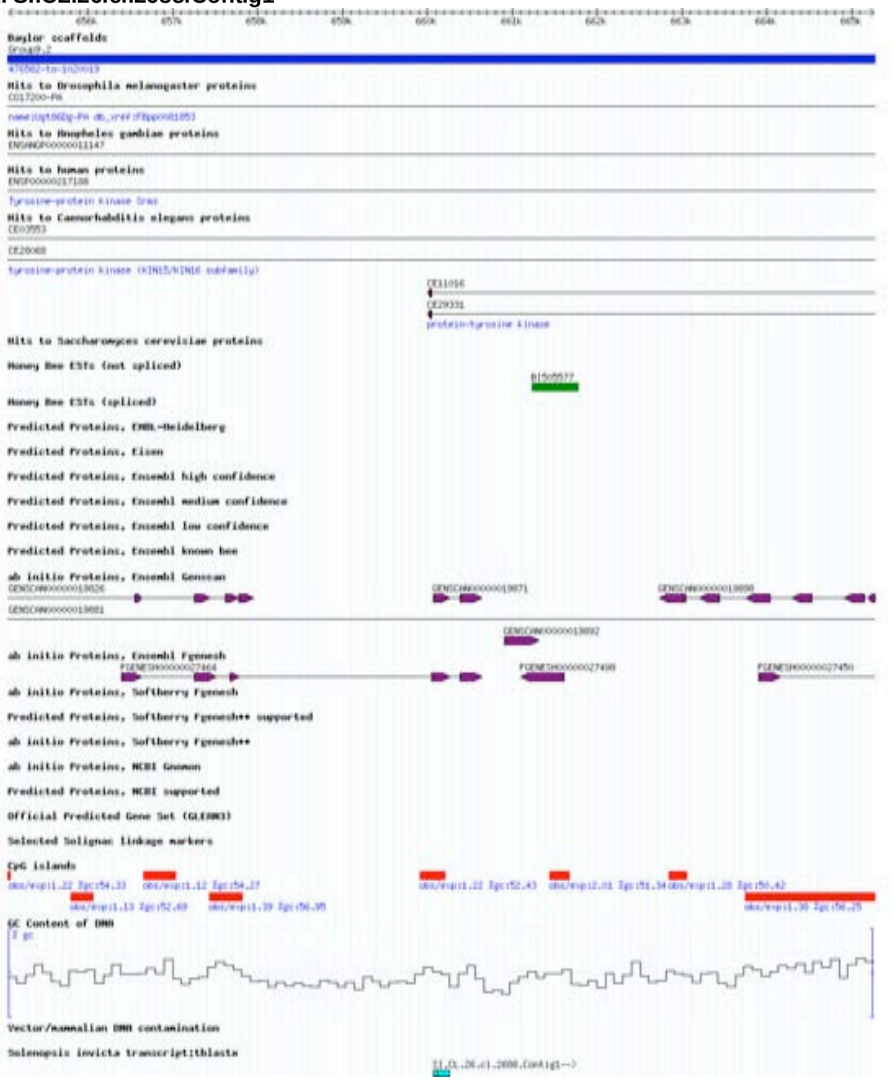
i.SI.CL.19.cl.1953.SiJWC11BBX.scf



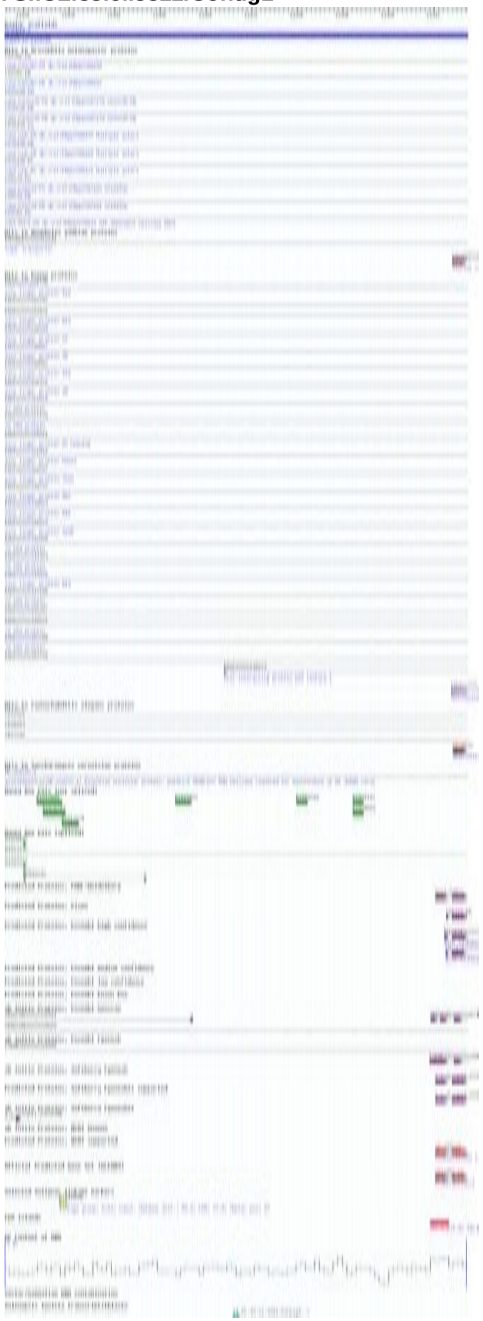
j. SI.CL.23.cl.2326.Contig1



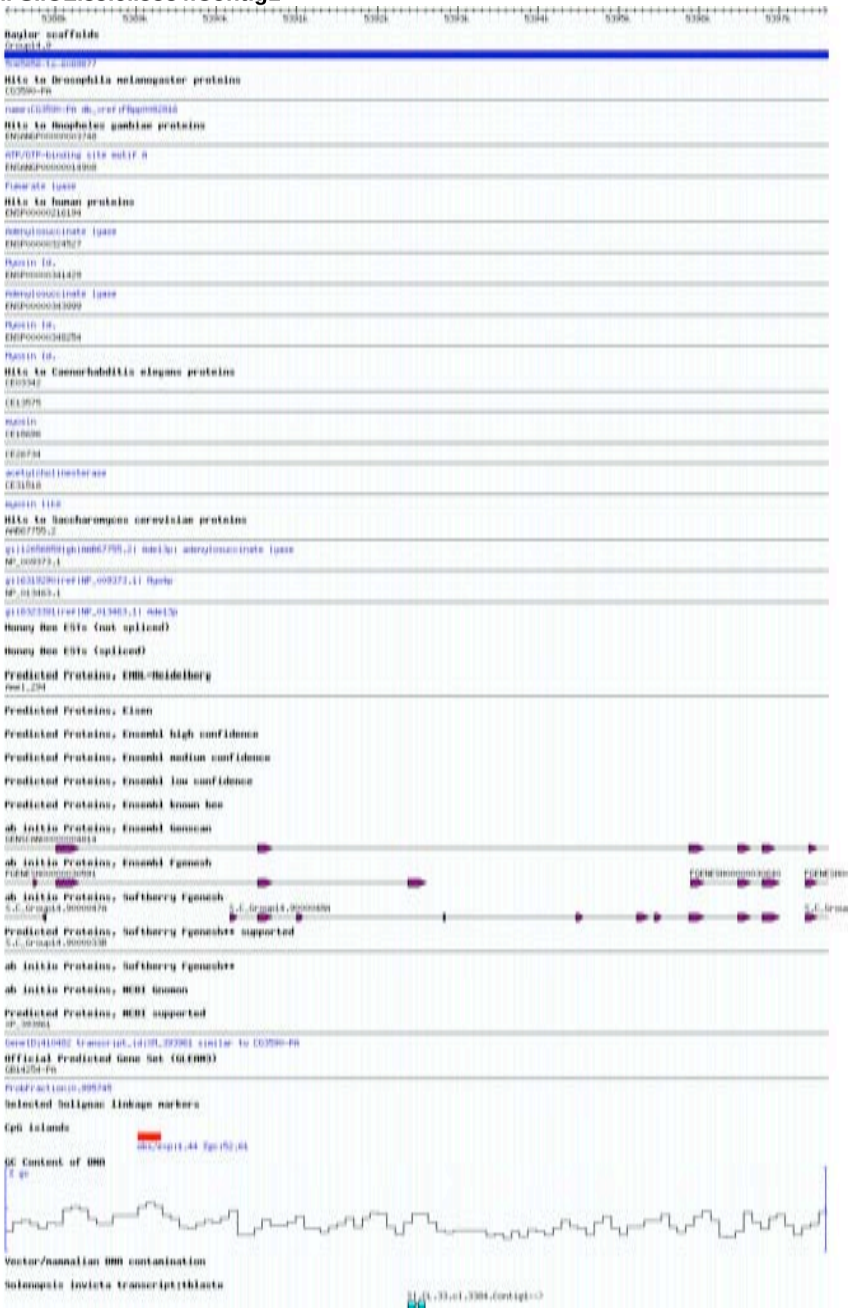
k. SI.CL.26.cl.2688.Contig1



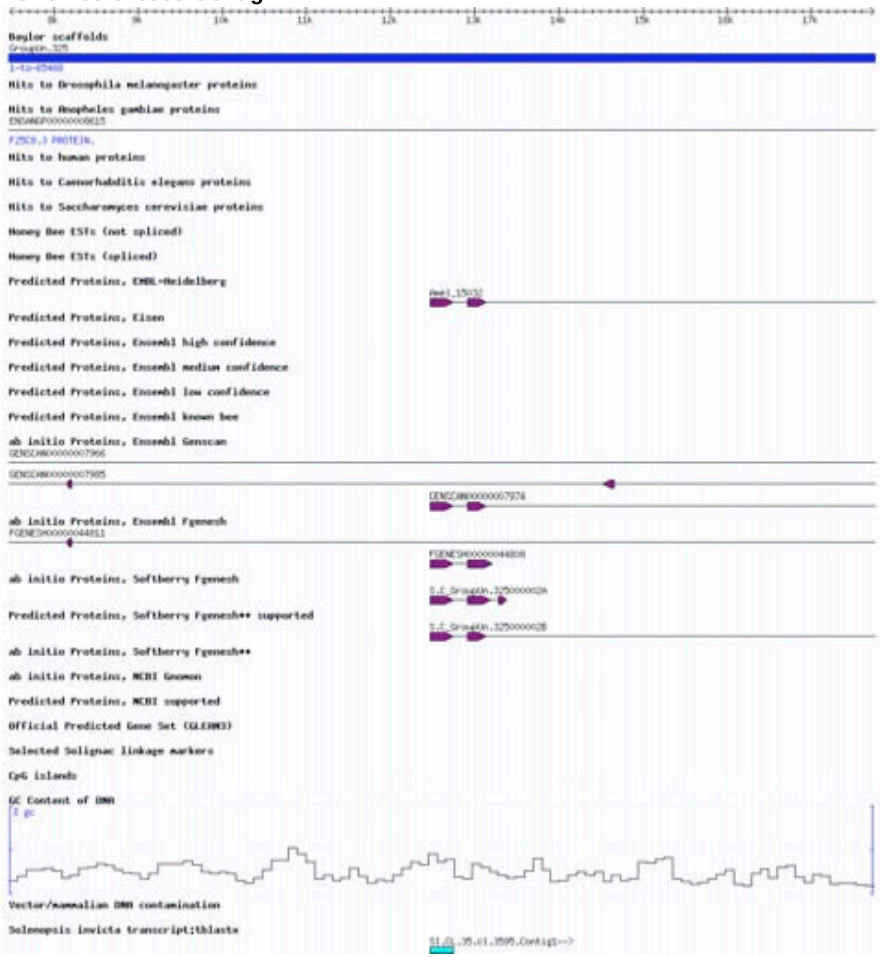
l. SI.CL.33.cl.3311.Contig1



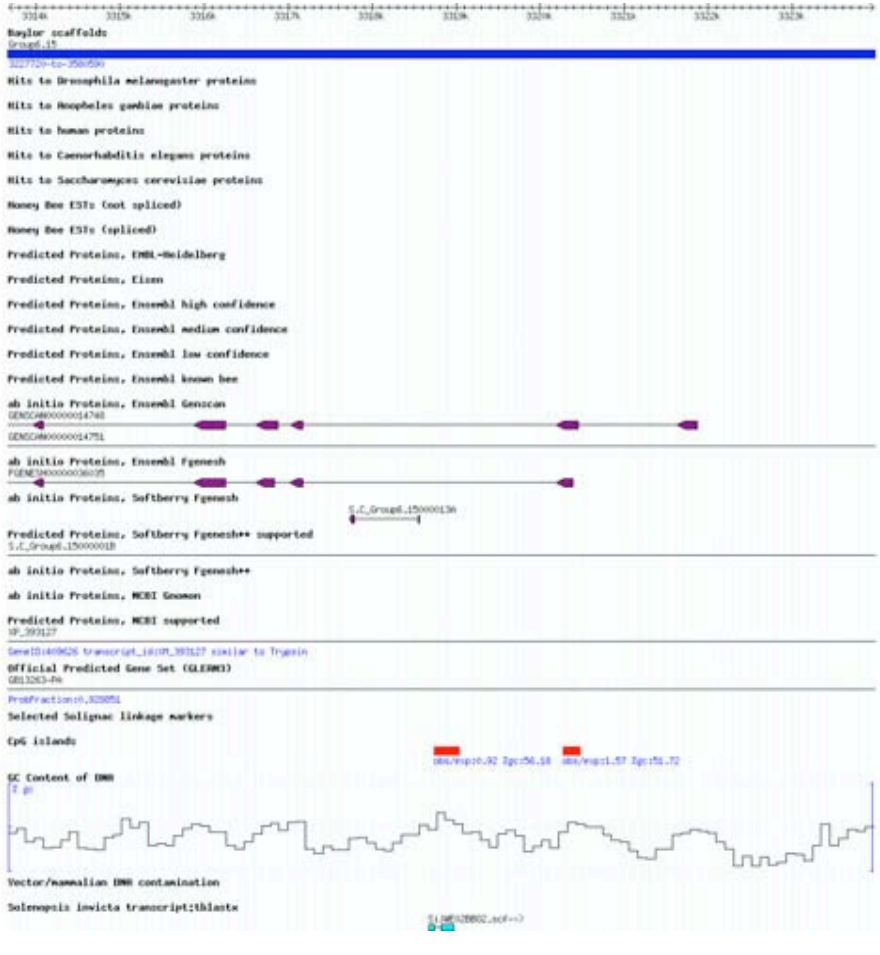
m. SI.CL.33.cl.3384.Contig1



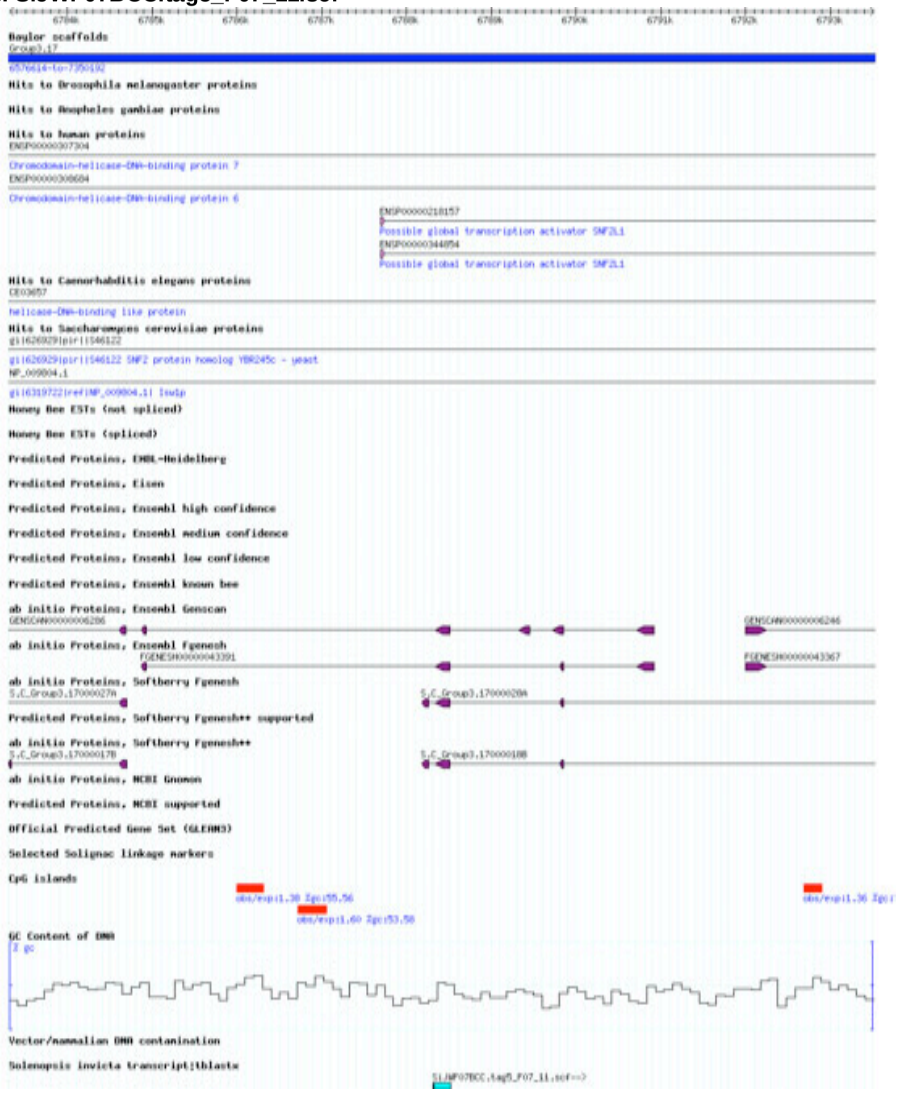
n. SI.CL.35.cl.3595.Contig1



o.SiJWE02BBO2.scf



p. SiJWF07BCC.tag5_F07_11.scf

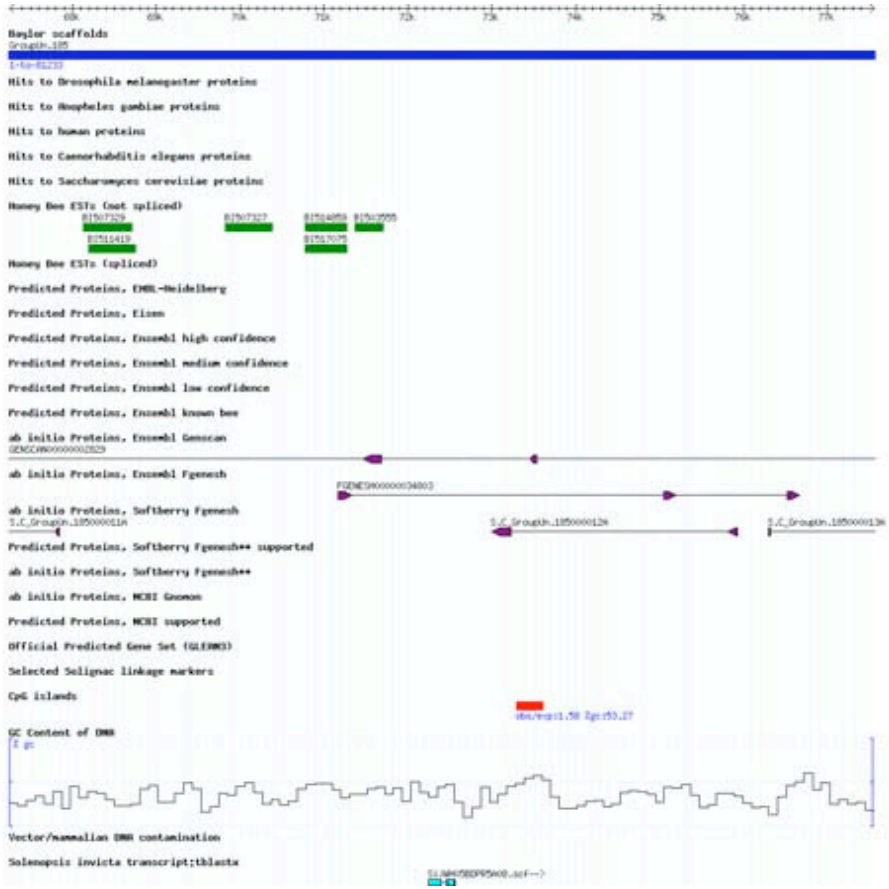


q.SiJWG01BDU2.scf

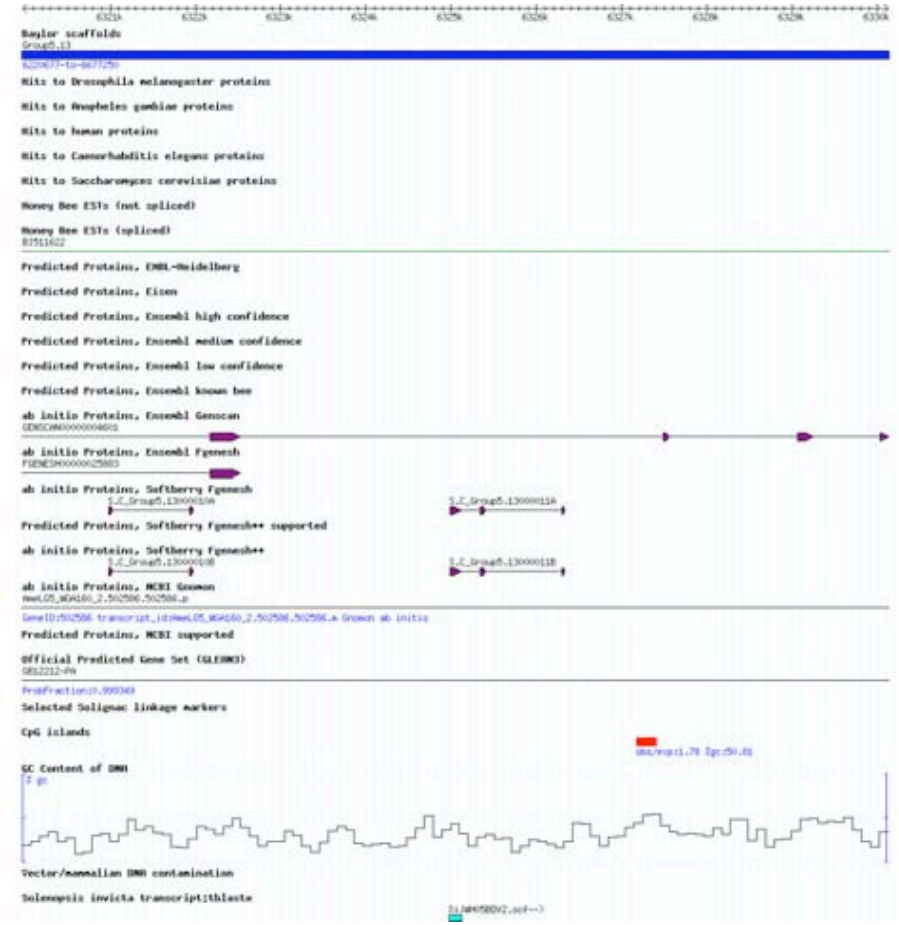
r. SiJWG03ACB.scf

s. SiJWH02AAN.scf

t. SiJWH05BDPR5A08.scf

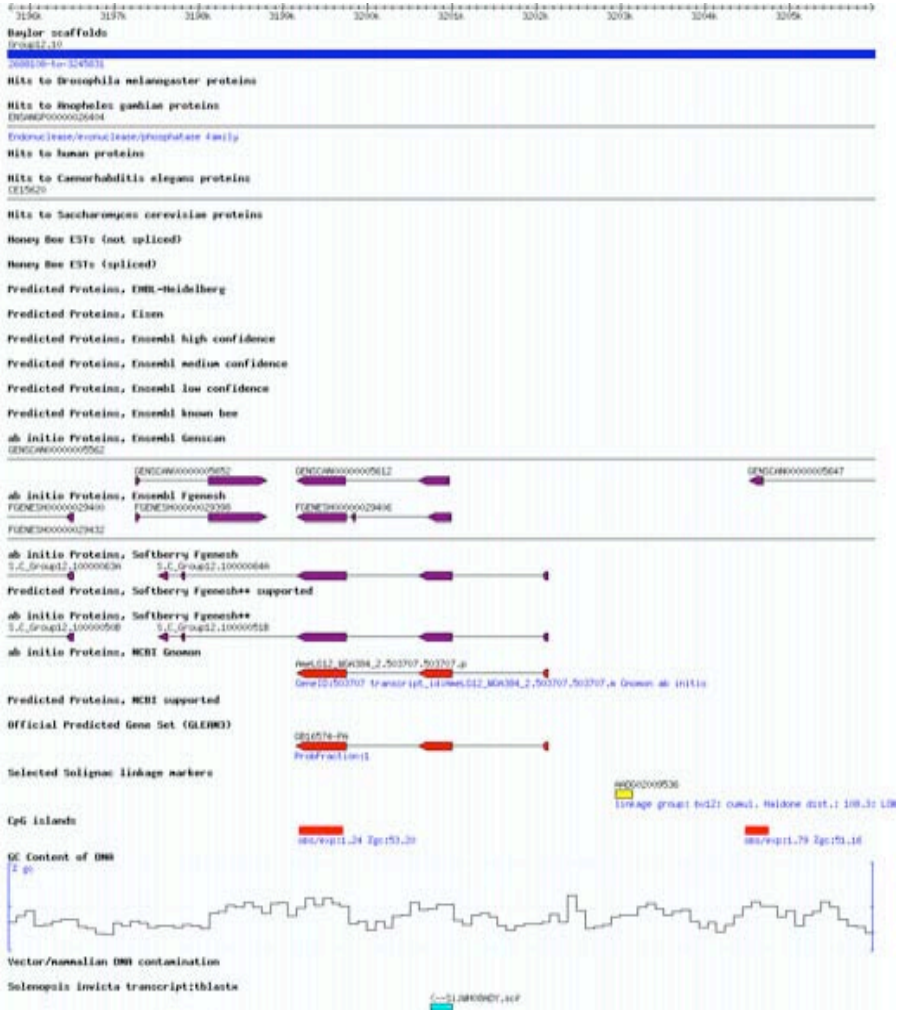


u. SiJWH05BDV2.scf



v.SiJWH08AAT.scf

w.SiJWH08ADY



Additional data file 6 Fire ant assembled sequences similar to D. melanogaster genes with the GO term 'behavior'

Additional data file 6: Fire ant assembled sequences similar to D. melanogaster genes annotated with behavior in any GO term.

There are 108 fire ant assembled sequences that match 81 unique Drosophila melanogaster genes

Fire ant assembled sequence	Gene name	E-value	Drosophila polypeptide ID
SI.CL.5.cl.511.Contig1	ATPsyn-beta-PA	1.00E-147	CG11154-PA
SiJWA03AAE.tag5_A03_01.scf	CG10724-PB	1.00E-125	CG10724-PB
SI.CL.34.cl.3440.Contig1	Abl-PA	1.00E-116	CG4032-PA
SI.CL.41.cl.4169.Contig1	drk-PB	1.00E-108	CG6033-PB
SI.CL.13.cl.1367.Contig1	CkIIbeta-PA	1.00E-104	CG15224-PA
SI.CL.0.cl.078.Contig1	Hsp83-PA	6.00E-97	CG1242-PA
SiJWG05CAC2.scf	tun-PE	9.00E-95	CG30084-PE
SI.CL.4.cl.457.Contig1	Gdh-PE	8.00E-93	CG5320-PE
SiJWD08BAP.scf	shot-PB	1.00E-88	CG18076-PB
SI.CL.12.cl.1279.Contig1	Vha55-PB	1.00E-87	CG17369-PB
SI.CL.6.cl.653.Contig1	bt-PE	1.00E-85	CG32019-PE
SI.CL.28.cl.2857.Contig1	Cdc42-PB	4.00E-85	CG12530-PB
SI.CL.27.cl.2793.Contig1	Ras85D-PA	4.00E-81	CG9375-PA
SiJWH07BBM2.scf	Ced-12-PA	7.00E-80	CG5336-PA
SiJWG03ACG.scf	betaTub56D-PB	2.00E-79	CG9277-PB
SI.CL.26.cl.2682.Contig1	sqh-PA	8.00E-79	CG3595-PA
SI.CL.25.cl.2532.Contig1	Rho1-PE	3.00E-74	CG8416-PE
SI.CL.33.cl.3317.Contig1	loj-PB	6.00E-74	CG10733-PB
SI.CL.13.cl.1344.SiJWC08BDJ.scf	crl-PA	1.00E-73	CG4443-PA
SI.CL.13.cl.1344.Contig1	crl-PA	5.00E-73	CG4443-PA
SI.CL.31.cl.3117.Contig1	tsr-PA	3.00E-72	CG4254-PA
SI.CL.4.cl.492.Contig1	awd-PA	1.00E-70	CG2210-PA
SI.CL.8.cl.878.Contig1	eIF-5C-PB	2.00E-68	CG2922-PB
SI.CL.2.cl.258.Contig1	TpnC73F-PA	5.00E-68	CG7930-PA
SiJWE02ABO.scf	Pka-R1-PG	4.00E-66	CG3263-PG
SiJWH01ACM.scf	slmo-PA	2.00E-65	CG9131-PA
SI.CL.9.cl.919.Contig1	betaTub56D-PB	9.00E-65	CG9277-PB
SiJWF08BAQ.scf	Vinc-PA	3.00E-60	CG3299-PA
SiJWG07BCO2.scf	Arc-p20-PA	5.00E-59	CG5972-PA
SiJWC12ABF.scf	sqd-PA	6.00E-59	CG16901-PA
SI.CL.15.cl.1563.Contig1	14-3-3epsilon-PD	1.00E-58	CG31196-PD
SI.CL.26.cl.2604.Contig1	lola-PS	3.00E-57	CG12052-PS
SI.CL.12.cl.1252.Contig1	Pax-PF	2.00E-56	CG31794-PF
SiJWG10ADB.scf	Pvr-PA	3.00E-56	CG8222-PA
SI.CL.12.cl.1285.Contig1	betaTub60D-PA	6.00E-55	CG3401-PA
SI.CL.6.cl.643.Contig1	Hmgcr-PB	2.00E-52	CG10367-PB
SI.CL.29.cl.2999.Contig1	p16-ARC-PA	8.00E-52	CG9881-PA
SiJWE08BAJ.scf	twf-PA	7.00E-50	CG3172-PA
SI.CL.4.cl.492.Contig2	awd-PA	3.00E-48	CG2210-PA
SI.CL.3.cl.390.Contig1	eIF-5C-PB	3.00E-44	CG2922-PB
SiJWE02ADO.scf	eas-PB	3.00E-44	CG3525-PB
SiJWE06ACV.scf	beta4GalNAcTA-PA	8.00E-44	CG8536-PA
SiJWD09ABM.scf	gammaTub23C-PA	6.00E-43	CG3157-PA
SiJWB06ABV.scf	Pka-C1-PB	2.00E-42	CG4379-PB
SI.CL.11.cl.1166.SiJWG04ADU.scf	robl-PA	1.00E-41	CG10751-PA
SI.CL.11.cl.1166.SiJWG07BBE.scf	robl-PA	1.00E-41	CG10751-PA
SI.CL.11.cl.1166.SiJWF02CAT2.scf	robl-PA	2.00E-41	CG10751-PA
SiJWD12ADC.scf	alphaTub84D-PA	2.00E-41	CG2512-PA
SI.CL.11.cl.1166.Contig2	robl-PA	2.00E-41	CG10751-PA
SI.CL.11.cl.1166.Contig1	robl-PA	2.00E-41	CG10751-PA
SI.CL.11.cl.1166.SiJWB11ACH.scf	robl-PA	1.00E-40	CG10751-PA
SI.CL.11.cl.1166.SiJWC07BDI2.scf	robl-PA	2.00E-40	CG10751-PA
SI.CL.42.cl.4217.Contig1	klg-PA	2.00E-40	CG6669-PA
SiJWH05AEA.scf	dco-PC	4.00E-40	CG2048-PC
SI.CL.17.cl.1743.Contig1	aay-PA	3.00E-39	CG3705-PA
SiJWB04BCC.tag5_B04_04.scf	Lar-PA	1.00E-38	CG10443-PA
SiJWH04ABK.scf	sesB-PA	2.00E-38	CG16944-PA
SiJWA04BDS2.scf	e-PA	7.00E-38	CG3331-PA
SI.CL.24.cl.2496.Contig1	sbr-PA	5.00E-37	CG1664-PA
SiJWF06ADX.scf	Gad1-PB	6.00E-37	CG14994-PB
SI.CL.40.cl.4022.Contig1	Dscam-PB	6.00E-37	CG17800-PB
SI.CL.27.cl.2741.Contig1	syt-PD	4.00E-36	CG3139-PD
SiJWG08BAP.scf	Ptp52F-PA	1.00E-35	CG18243-PA
SiJWC05ACO.scf	Gas8-PB	9.00E-35	CG14271-PB
SiJWH08CAV.scf	bt-PE	1.00E-34	CG32019-PE
SI.CL.21.cl.2105.Contig1	TpnC25D-PA	9.00E-34	CG6514-PA
SiJWB01BCB2.scf	rok-PA	1.00E-33	CG9774-PA
SI.CL.19.cl.1905.Contig1	Tsp74F-PA	5.00E-30	CG5492-PA
SiJWG03BAC.scf	stck-PA	5.00E-30	CG7954-PA
SI.CL.8.cl.865.Contig1	Egfr-PB	7.00E-30	CG10079-PB
SI.CL.23.cl.2317.Contig1	alphaTub84D-PA	4.00E-28	CG2512-PA
SiJWA02AAP.scf	Tsp74F-PA	3.00E-27	CG5492-PA
SI.CL.0.cl.091.Contig1	LanA-PA	4.00E-27	CG10236-PA
SiJWB09ADK.scf	CG3599-PA	6.00E-25	CG3599-PA
SiJWG01ADR.scf	mnb-PC	1.00E-24	CG7826-PC
SI.CL.4.cl.440.Contig1	Tsp74F-PA	2.00E-24	CG5492-PA
SiJWH06BDP.scf	tko-PB	2.00E-24	CG7925-PB
SiJWE09BCU2.scf	bt-PE	3.00E-22	CG32019-PE
SI.CL.25.cl.2558.Contig1	ATPsyn-beta-PA	4.00E-21	CG11154-PA
SiJWB08BBS.scf	alphaTub84D-PA	3.00E-17	CG2512-PA
SI.CL.42.cl.4217.SiJWB52C03_pcr1_ta	klg-PA	8.00E-16	CG6669-PA
SI.CL.3.cl.321.Contig2	Gad1-PB	3.00E-15	CG14994-PB
SI.CL.42.cl.4217.SiJWB52C04_pcr1_ta	klg-PA	1.00E-14	CG6669-PA
SI.CL.39.cl.3947.Contig1	dock-PB	1.00E-13	CG3727-PB
SiJWF02BDZ.scf	Hk-PA	1.00E-13	CG32688-PA
SiJWA07AAS.scf	ninA-PA	1.00E-12	CG6449-PA
SiJWC07ACO.scf	lola-PB	1.00E-12	CG12052-PB
SI.CL.20.cl.2042.Contig1	tun-PE	1.00E-12	CG30084-PE
SiJWB10ADI.scf	lola-PB	2.00E-12	CG12052-PB
SI.CL.1.cl.185.Contig1	sli-PC	6.00E-12	CG8355-PC
SiJWB11ABH.scf	for-PG	1.00E-11	CG10033-PG
SiJWB03ACL.scf	CadN-PH	2.00E-11	CG7100-PH
SI.CL.10.cl.1053.Contig1	lola-PU	4.00E-10	CG12052-PU
SI.CL.4.cl.485.Contig3	spen-PA	1.00E-09	CG18497-PA
SI.CL.24.cl.2486.Contig1	slbo-PA	2.00E-09	CG4354-PA
SiJWF03BDD2.scf	Mhcl-PE	3.00E-09	CG31045-PE
SiJWE03CAX.scf	vanin-like-PA	2.00E-08	CG32754-PA
SiJWF04BBP.scf	gish-PE	8.00E-08	CG6963-PE
SiJWD03ACB.scf	Ddc-PA	1.00E-07	CG10697-PA
SiJWB02ABH.scf	MICAL-PC	1.00E-07	CG33208-PC
SI.CL.34.cl.3448.Contig1	sbr-PA	2.00E-07	CG1664-PA
SI.CL.31.cl.3175.Contig1	BicC-PD	2.00E-07	CG4824-PD
SI.CL.12.cl.1287.SiJWH09BBY.scf	th-PC	3.00E-07	CG12284-PC
SiJWB07BBZ.scf	zip-PD	4.00E-07	CG15792-PD
SI.CL.28.cl.2854.Contig1	Egfr-PB	6.00E-07	CG10079-PB
SI.CL.30.cl.3038.Contig1	Sema-5c-PA	9.00E-07	CG5661-PA
SI.CL.25.cl.2515.Contig1	Dhc36C-PA	2.00E-06	CG5526-PA
SiJWC09ACE.scf	lola-PU	5.00E-06	CG12052-PU

Additional data file 7 Annotated list of the most abundant transcripts

Fire Ant Contig Name	Number of ESTs	ID of Best Blastx Hit	Hit Description [Organism with Best Hit]	E-value	Identity (%)
SI.CL.0.cl.071.Contig1	48	ENSDARP00000052130	Conserved protein of unknown function [Danio rerio]	2.00E-17	79.2
SI.CL.4.cl.464.Contig2	39	Q6PWE1	Cytochrome b [Formica lugubris]	1.00E-113	60.1
SI.CL.1.cl.121.Contig2	38	P29875	Cytochrome c oxidase subunit 2 [Lasius sp]	2.00E-57	55.4
SI.CL.0.cl.037.Contig1	33	Q6IT30	Weak hit to NADH dehydrogenase subunit 3 [Philaenus spumarius]	8.00E-05	76.9
SI.CL.0.cl.051.Contig1	26	no hit	Putative 16S rRNA, mitochondrial (according to Blastn)	na	na
SI.CL.1.cl.175.Contig2	24	Q5JCK9	NADH dehydrogenase subunit 2 [Aedes albopictus]	1.00E-07	30.9
SI.CL.0.cl.005.Contig1	17	ENSAPMP00000034158	Putative chymotrypsin [Apis mellifera]	7.00E-51	43.9
SI.CL.4.cl.485.Contig3	17	Q5CEG7	Senescence-associated protein [Cryptosporidium hominis]	3.00E-34	70.2
SI.CL.7.cl.734.Contig1	16	O61334	Elongation factor-1alpha F2 [Apis mellifera]	1.00E-112	97.1
SI.CL.0.cl.065.Contig5	15	ENSCBRP00000009337	Hypothetical protein [Caenorhabditis briggsae]	1.00E-23	43.4
SI.CL.1.cl.159.Contig1	15	Q7YXM2	Putative thioredoxin peroxidase [Apis mellifera ligustica]	1.00E-82	87.5
SI.CL.17.cl.1723.Contig1	15	ENSAPMP00000012354	Putative 60S ribosomal protein L24 [Apis mellifera]	1.00E-51	69.5
SI.CL.8.cl.847.Contig1	15	Q29CT8	Putative NADH dehydrogenase 1 alpha subcomplex 4 [Drosophila pseudoobscura]	8.00E-23	61.7
SI.CL.0.cl.052.Contig1	14	Q964Q9	MRNA cap-binding protein eIF4E [Spodoptera frugiperda]	1.00E-71	61.4
SI.CL.1.cl.194.Contig3	14	ENSAPMP00000015055	Putative ribosomal protein Ubq/L40e [Apis mellifera]	3.00E-64	100.0
SI.CL.11.cl.1121.Contig1	14	Q71U45	Cytochrome c oxidase VIIc [Homo sapiens]	6.00E-11	55.6
SI.CL.6.cl.610.Contig1	14	Q8QY61	Polyprotein [Sacbrood virus]	2.00E-11	26.5
SI.CL.7.cl.796.Contig1	14	ENSAPMP00000002127	Putative ribosomal protein L36 [Apis mellifera]	3.00E-53	90.4
SI.CL.1.cl.195.Contig1	13	Q3LW33	Weak hit to hypothetical protein [Chlorarachnion sp]	2.5	26.2
SI.CL.15.cl.1507.Contig1	13	ENSAPMP00000035247	Similar to boca [Apis mellifera]	4.00E-27	48.9
SI.CL.16.cl.1664.Contig1	13	ENSAPMP00000000231	Putative ribosomal protein L26 [Apis mellifera]	2.00E-51	75.0
SI.CL.2.cl.276.Contig1	13	SINFRUP00000147381	Weak hit to hypothetical protein [Takifugu rubripes]	0.002	25.7
SI.CL.2.cl.287.Contig1	13	Q9V754	CG10153-PA unknown protein [Drosophila melanogaster]	3.00E-75	70.8
SI.CL.4.cl.493.Contig1	13	Q4PP87	3-hydroxyacyl-CoA dehydrogenase [Lysiphlebus testaceipes]	1.00E-101	74.9
SI.CL.5.cl.569.Contig1	13	Q29MZ9	GA11014-PA (Fragment) [Drosophila pseudoobscura]	1.00E-14	40.8
SI.CL.5.cl.592.Contig2	13	Q4DNK9	Weak hit to hypothetical protein [Trypanosoma cruzi]	4.5	32.7
SI.CL.0.cl.016.Contig1	12	ENSAPMP00000024841	Putative Replication factor C (Activator 1) 3 [Apis mellifera]	2.33E-156	78.3
SI.CL.1.cl.165.Contig2	12	ENSAPMP00000012348	Putative mitochondrial thioredoxin 2 [Apis mellifera]	2.00E-57	78.4
SI.CL.13.cl.1316.Contig1	12	ENSAPMP00000010338	Putative mitochondrial ATP synthase O subunit [Apis mellifera]	2.00E-70	65.0
SI.CL.2.cl.230.Contig1	12	ENSAPMP00000018003	Similar to enhancer of rudimentary [Apis mellifera]	2.00E-19	45.4
SI.CL.2.cl.237.Contig1	12	Q23AX3	Weak hit to chlamydial polymorphic outer membrane protein repeat containing prote	0.56	27.7
SI.CL.3.cl.337.Contig1	12	ENSAPMP00000001195	Putative ribosomal protein L15 [Apis mellifera]	1.00E-89	82.8
SI.CL.5.cl.508.Contig1	12	ENSAPMP00000013220	Putative chymotrypsinogen [Apis mellifera]	3.00E-40	38.7
SI.CL.8.cl.853.Contig1	12	ENSAPMP00000005757	Putative cytochrome c [Apis mellifera]	7.00E-53	86.1
SI.CL.9.cl.938.Contig1	12	Q9VKE6	CG14934-PA [Drosophila melanogaster]	1.00E-140	51.5
SI.CL.9.cl.985.Contig1	12	Q7RP30	Weak hit to hypothetical protein [Plasmodium yoelii yoelii]	2.4	29.5
SI.CL.0.cl.038.Contig1	11	Q7Z161	Ferritin [Leptinotarsa decemlineata]	2.00E-52	51.2
SI.CL.0.cl.041.Contig1	11	Q85H57	Cytochrome c oxidase subunit I (Fragment) [Myrmica ruginodis]	1.00E-175	66.7
SI.CL.10.cl.1001.Contig2	11	ENSAPMP00000016467	Putative translation initiation factor 4A [Apis mellifera]	0	92.2
SI.CL.11.cl.1190.Contig1	11	ENSAPMP00000029078	Putative unknown protein [Apis mellifera]	1.00E-39	52.0
SI.CL.12.cl.1246.Contig1	11	ENSAPMP00000024992	Putative Acyl-CoA-binding protein homolog (Diazepam-binding inhibitor homolog) [A	3.00E-25	62.9
SI.CL.13.cl.1304.Contig1	11	Q9NH75	Venom allergen 4 precursor [Solenopsis geminata]	3.00E-43	57.7
SI.CL.2.cl.241.Contig1	11	Q7Q5H5	Chitin binding Peritrophin-A domain containing protein [Anopheles gambiae str]	1.00E-09	40.3
SI.CL.4.cl.454.Contig1	11	ENSAPMP00000027689	Putative lectin 4 C-type lectin [Apis mellifera]	1.00E-40	88.2
SI.CL.5.cl.510.Contig2	11	Q6PCI5	Weak hit to Mtmr3 protein [Xenopus laevis]	1	43.1
SI.CL.5.cl.580.Contig1	11	Q4GX85	Ribosomal protein L39e [Biphyllus lunatus]	4.00E-21	100.0
SI.CL.5.cl.592.Contig4	11	ENSAPMP00000014490	Weak hit to putative unknown protein [Apis mellifera]	0.001	29.0
SI.CL.6.cl.625.Contig1	11	Q5AJE2	Weak hit to hypothetical protein RNA1 [Candida albicans SC5314]	2.3	38.9
SI.CL.7.cl.776.Contig2	11	Q4UJ03	Weak hit to hypothetical protein [Theileria annulata]	3.5	27.6
SI.CL.8.cl.820.Contig1	11	ENSAPMP00000032435	Putativechymotrypsin [Apis mellifera]	2.00E-58	47.7
SI.CL.8.cl.828.Contig1	11	Q6B863	Weak hit to cytochrome c oxidase polypeptide VIII [Ixodes pacificus]	1.2	36.6
SI.CL.0.cl.031.Contig1	10	O88668	CREG1 protein precursor (Cellular repressor of E1A-stimulated genes 1) [Mus musc	6.00E-37	43.8
SI.CL.0.cl.098.Contig1	10	Q56FH6	Ribosomal protein S7 [Lysiphlebus testaceipes]	3.00E-80	79.3
SI.CL.1.cl.148.Contig1	10	Q7P4X3	Weak hit to GTP cyclohydrolase I [Fusobacterium nucleatum subsp]	5.5	22.7
SI.CL.1.cl.182.Contig1	10	Q2VW29	Antennae-specific chemosensory protein [Solenopsis invicta]	5.00E-20	43.4
SI.CL.10.cl.1023.Contig1	10	Q5EN85	Weak hit to oxidative stress protein [Aurelia aurita]	0.005	32.8
SI.CL.12.cl.1201.Contig1	10	Q8E2N6	Weak hit to hypothetical protein gbs2086 [Streptococcus agalactiae serotype III]	3.5	30.6
SI.CL.14.cl.1425.Contig1	10	Q4XME0	Weak hit to hypothetical protein (Fragment) [Plasmodium chabaudi]	0.94	27.5
SI.CL.14.cl.1484.Contig1	10	ENSAPMP00000034161	Weak hit to putative unknown protein [Apis mellifera]	0.038	56.3
SI.CL.18.cl.1828.Contig1	10	ENSAPMP00000019311	Putative ribosomal protein S13 [Apis mellifera]	1.00E-78	96.0
SI.CL.2.cl.243.Contig1	10	Q4I318	Weak hit to hypothetical protein [Gibberella zeae]	3.5	23.9
SI.CL.2.cl.271.Contig1	10	Q2F5V9	Signal sequence receptor beta subunit [Bombyx mori]	5.00E-59	66.5
SI.CL.3.cl.331.Contig4	10	no hit	Putative 12S rRNA, mitochondrial (according to Blastn)	na	na
SI.CL.4.cl.468.Contig1	10	ENSAPMP00000004807	Putative ribosomal protein S29e [Apis mellifera]	2.00E-27	98.2
SI.CL.6.cl.634.Contig1	10	ENSAPMP00000019417	Weak hit to putative unknown protein [Apis mellifera]	1.2	27.0
SI.CL.8.cl.829.Contig1	10	Q28BN5	NADH dehydrogenase (Ubiquinone) 1 alpha subcomplex 5 [Xenopus tropicalis]	1.00E-32	56.1
SI.CL.9.cl.927.Contig1	10	ENSAPMP00000031188	Similar to B-cell receptor-associated protein 31 [Apis mellifera]	2.00E-60	58.2

Additional data file 8 PCR results for the cDNA clones deposited onto the microarray

Additional data file 9 Fire ant assembled sequences that had at least one cDNA clone with a good (single-band) PCR product

Additional data file 10 Details on the microarray analyses performed

Additional data file 10

Evaluation of the fire ant microarray

To evaluate the percentage of cDNA spots derived from legitimate and sufficiently highly expressed transcripts, we examined the signal-to-background value of all spots in four test hybridizations. These four hybridizations were technical replicates of two biological samples.

Biological Samples:

Brood: eggs, larvae and pupae of all castes in equal amounts
Adults: workers, virgin queens, and males from both colony types in equal amounts

Total RNA was amplified using the MessageAmp II kit (Ambion). “Brood” RNA was independently amplified three times while “Adult” RNA was independently amplified two times. Amplified RNA was labeled using the “indirect” method where reverse transcription was performed in the presence of amino-allyl dUTP and the resulting cDNA subsequently coupled to Cy3 or Cy5 fluorescent monomers. The samples were labeled and hybridized onto the microarrays as outlined below.

For each microarray, we counted the number of spots that were derived from a single good PCR product and that had a signal intensity greater than background plus two standard deviations for either the Cy3 or Cy5 labeled samples. Using the conservative criterion that a spot has sufficiently intense signal on all four microarrays, 82.8% (14,642/17,685) were satisfactory. On average 93.8% (16,588/17,685) had an adequate signal. These data indicate that most cDNA clones are derived from legitimate transcripts.

Supplementary Table Microarray: Hybridization schematic and the number of spots with satisfactory signal for each microarray

		Sample*	
Slide ID	Cy3-labeled	Cy5-labeled	Spots with adequate signal
B109	adult-1	brood-1	16,203
B110	brood-2	adult-1	16,534
C021	adult-2	brood-3	16,794
C022	brood-3	adult-2	16,822
Total spots on microarray (1 pcr product)			17,685
Average of 4 microarrays			16,588.25 (93.8%)
All hybridizations			14,642 (82.8%)

* Numbers refer to different amplifications from same initial total RNA

Adult and Brood enriched genes

We generated lists of potentially adult- or brood-enriched genes using two methods. In the first method we used a simple 4-fold expression difference cutoff (Additional data file 11). In the second method we performed a 1-sample t-test using a p-value cutoff of 0.001 (Additional data file 12). For both analyses, data were first normalized (loess normalization by print-tip group) using the marray package of Bioconductor (R).

We have not performed additional statistical analyses on this data set as these microarray data are derived from single biological samples.

Additional data file 11 Fire ant clones that are differentially expressed between adults and brood based on a 4-fold cutoff

Additional data file 12 Fire ant clones that are differentially expressed between adults and brood based on a t-test ($p < 0.001$)

Part III

GENOMIC INVESTIGATIONS OF ANTS

5

CHANGES IN REPRODUCTIVE ROLES ARE ASSOCIATED WITH WIDESPREAD CHANGES IN GENE EXPRESSION IN FIRE ANTS

by Yannick Wurm, John Wang and Laurent Keller

I performed all experiments and analyses and wrote the manuscript. John Wang and Laurent Keller helped design the experiment and revised the manuscript.

ABSTRACT

In many animals living in groups the reproductive status of individuals is determined by their social status. In species with social hierarchies, the death of dominant individuals typically upheaves the social hierarchy and provides an opportunity for subordinate individuals to improve their social status. Such a phenomenon occurs in the monogyne form of the fire ant, *Solenopsis invicta*, where colonies typically contain a single wingless reproductive queen, thousands of workers and hundreds of winged non-reproductive virgin queens. Upon the death of the mother queen, many virgin queens shed their wings and initiate reproductive development instead of departing on a mating flight. Workers progressively execute almost all of them over the following weeks. The workers base their collective decision on pheromonal cues associated with the onset of reproductive development of the virgin queens which occurs after orphaning.

To identify the molecular and genetic mechanisms behind the onset of competition between virgin queens, we collected virgin queens before the loss of their mother queen, six hours after orphaning and 24 hours after orphaning. Their RNA was extracted and hybridized against microarrays to examine the expression levels of approximately 10,000 genes. We identified 297 genes that were consistently differentially expressed after orphaning. These include genes that are putatively involved in the signaling and onset of reproductive development, as well as genes underlying major physiological changes in the young queens.

This article is in review for publication in *Molecular Ecology*

5.1 BACKGROUND

Reproduction is monopolized by only a small number of individuals in many group-living animals. Which individuals reproduce can be determined by fights for dominance or territory, by seniority within the group, by genotype and by other factors (Solomon & French 1997; Keller 1993; Keller & Ross 1998; Keller & Reeve 1994). Although the social stimuli responsible for changes in reproductive hierarchies are well-documented in many animals (Solomon & French 1997), only a few studies have examined the molecular and physiological mechanisms linking such stimuli to changes in reproductive status. In the cichlid fish *Astatotilapia burtoni*, disappearance of the dominant male leads to rapid reactions in subordinate males, including dramatic changes in body coloration and behavior, growth of certain brain regions and increases in brain levels of *gonadotropin releasing hormone 1* and *early growth response factor 1* (Burmeister *et al.* 2005; White *et al.* 2002). Similarly, the transition from subordinate to breeder status in white-browed sparrow weavers is accompanied by changes in type of song, morphology of song-related brain areas, and an increase in levels of two hormone receptors and two synaptic proteins in a song-related brain area (Voigt *et al.* 2007). Changes in brain morphology also accompany the transition from subordinate to breeder status in naked mole rats (Holmes *et al.* 2007). While the previous studies provide valuable insight into the responses to social opportunities, they mostly focused on brain morphology and only few candidate genes.

Social insects provide excellent models for studying the mechanisms involved in reproductive competition (Roseler 1991; Roseler *et al.* 1984; Dietemann *et al.* 2006; Neumann *et al.* 2000; Keller 1993). In social bees, wasps and ants there is a clear division of labor with one or a few individuals monopolizing reproduction. Differences in reproductive roles are generally associated with tremendous physiological and behavioral modifications (Wilson 1971; Bourke & Franks 1995). This has led to many behavioral and hormone-based experiments including some in fire ants that have even succeeded in isolating glands and compounds involved in maintaining social dominance hierarchies (Vargo & Laurel 1994; Vargo & Hulsey 2000; Vargo 1999; Brent & Vargo 2003; Vander Meer *et al.* 1980). Investigating social life at a larger scale has only recently become possible with the development of genomic tools for social insects (The Honey Bee Genome Sequencing Consortium 2006; Wang *et al.* 2007; Wurm *et al.* 2009). Some of the first studies focused on identifying the genes involved in differences between reproductive and non-reproductive castes (Pereboom *et al.* 2005; Weil *et al.* 2009; Gräff *et al.* 2007), and others have investigated the link between social context and gene activity (Toth *et al.* 2007; Wang *et al.* 2008). However, very little still is known about the changes in gene expression associated with changes in reproductive roles.

The red imported fire ant, *Solenopsis invicta,* represents a particularly attractive model for studying the onset of competition between subordinate individuals. During the reproductive season, colonies of the monogyne form (single queen per colony) can produce hundreds or even thousands of young virgin daughter queens. These virgin queens spend the next few weeks building up fat reserves within the colony. Once they reach sexual maturity, they do not immediately become reproductive because such queens are perceived as threats to the dominant queen and executed by the workers (Vargo & Porter 1993; Vargo & Laurel 1994). Thus, virgin queens remain in the parental nest without reproducing until they participate in a mating flight and attempt to found their own colony. However, a remarkable alternative exists in *S. invicta* when the mother queen dies. During the days after orphaning, many young queens shed their wings and initiate reproductive development. This reproductive shift is accompanied by the emission of pheromonal signals to which nestmate queens and workers react. When virgin nestmate queens perceive such signals, they refrain from shedding their own wings and initiating reproductive development (Fletcher *et al.* 1983; Vargo 1999). When orphaned workers perceive pheromonal signals emitted by queens initiating reproductive development, they begin to tend to these queens (Fletcher & Blum 1981). However, if several queens produce signals associated with initiation of reproductive development, the workers will progressively execute almost all of them over the next few weeks (Fletcher

& Blum 1983). The surviving virgin queen or queens are thus "elected" by workers to replace the mother queen. These queens are unmated and thus unable to replenish the colony's worker force. However, until the colony's workers have died out, the queens can lay thousands of haploid eggs that develop into haploid reproductive males (Tschinkel 2006).

The aim of this study was to examine the molecular-physiological effects of the loss of the dominant mother queen on subordinate virgin queens that would compete for reproductive dominance. For this, we conducted orphaning simulations and examined gene expression using a microarray representing some 10,000 genes. We identified several categories of genes that are consistently upregulated after orphaning, some of which are also upregulated at the onset of reproductive development in other insects.

5.2 MATERIALS AND METHODS

Ant collection and rearing

Eight monogyne *S. invicta* fire ant colonies, each containing at least 50 winged virgin queens, were collected in Athens and Lexington, GA, USA in June 2006. All colonies were returned to the laboratory and reared for one month under standard conditions (Jouvenaz *et al.* 1977). Queen and male destined brood were removed weekly. We determined that each study colony was of the monogyne social form using several lines of evidence. Nest shape, nest density and worker size distribution were used to make initial identifications of social form in the field (Shoemaker *et al.* 2006). Subsequently, monogyny was confirmed for each colony by the presence of a single, highly physogastric, wingless queen. Finally, the social form was further verified by electrophoretically detecting only the *B* but not the *b* allele of *Gp-9* in pooled samples of 20 workers from each colony (lack of the *b* allele is diagnostic for monogyny in *S. invicta* in the USA (Ross 1997; Keller & Ross 1998; Krieger & Ross 2002; Shoemaker *et al.* 2006)).

Orphaning simulation, RNA isolation and microarray hybridization

We removed the mother queen and collected virgin queens just before orphaning as well as 6 and 24 hours after orphaning (subsequently referred to as time points t_{0h}, t_{6h} and t_{24h}) to examine the molecular reaction to orphaning in virgin *S. invicta* queens. However, virgin queens emit pheromonal signals after orphaning that are similar to those of an functional queen and can thus influence each other (Vargo 1999; Fletcher *et al.* 1983). We attempted to minimize such effects and simplify interpretation of results as follows: For t_{0h}, we haphazardly collected five virgin queens from the foraging area of each source colony and individually flash-froze them with liquid nitrogen in tubes containing 1g of 1.4mm Zirconium Silicate beads (QuackenBush). We placed ten additional virgin queens per source colony into individual colony fractions with 2g of mixed workers and brood. This simulated orphaning for a total of 80 virgin queens from a total of eight source colonies. We harvested half of the virgin queens thus treated after 6 hours (t_{6h}) and the remaining queens after 24 hours (t_{24h}). All collected queens were individually flash-frozen immediately after collection as described above. Samples were then stabilized until RNA isolation by the addition of 900 μl of cold Trizol reagent (Invitrogen) followed by homogenization with a FastPrep instrument (MP Biomedicals) and storage at -80°C.

Total RNA was isolated from all individuals sing the Trizol protocol. RNA was pooled from 5 individuals per source colony for each time point and treated with DNA-free (Ambion). Subsequently, impurities were filtered away with MicroCon-30 spin columns (Millipore), and RNA quality was assessed on a 1% agarose gel prior to amplification using the MessageAmp II kit (Ambion). Amplified mRNA samples from the eight colonies at three timepoints (t_{0h}, t_{6h} and t_{24h}) were labeled, hybridized to

microarrays made from 22,560 independent fire ant cDNA spots, and scanned as previously described (Wang *et al.* 2008). This was done according to a dye-balanced loop design (Supporting Figure 1). For all procedures, precautions including randomization of sample order were taken to avoid introducing unwanted biases.

Microarray analysis

Median signal and background levels for each probe were extracted from scanned microarray images using Axon Genepix software. The Bioconductor limma 2.16 package (Gentleman *et al.* 2004; Smyth 2004) in R 2.8.1 (The R Development Core Team 2007) was used for normexp background correction, print-tip loess normalization within arrays, and aquantile normalization between arrays. The arrayQualityMetrics package (Kauffmann *et al.* 2009) and custom R scripts were used for quality control. The 18,444 *Solenopsis invicta* cDNA spots yielding a single PCR band (Wang *et al.* 2007) and passing visual and automated inspection were used for analysis. The limma package was used to construct a design matrix incorporating the three sampling times (t_{0h}, t_{6h} and t_{24h}) and effects for colony and dye prior to bayesian fitting of the model. Differential expression was determined for the contrasts "t_{24h} vs. t_{0h}", "t_{24h} vs. t_{6h}", and "t_{6h} vs. t_{0h}" according to the nested F method in limma. Briefly, a moderated F test determined that 521 microarray clones were differentially expressed for at least one of the contrasts (with a 10% Benjamini-Hochberg False Discovery Rate). Subsequently, significance of differential expression was assigned to one or several contrasts.

Sequence data, annotation and gene category analysis

Microarray clone sequences (Wang *et al.* 2007) were combined with two runs of 454 cDNA sequencing (Y. Wurm, D. Hahn and DD. Shoemaker , unpublished; DH and DDS are at USDA-ARS, Gainesville). High quality sequence information was thus available for 16,227 out of the 18,444 *S. invicta* cDNA clones used for gene expression analysis. This was also the case for 475 out of the 521 significantly differentially expressed clones.

Annotation was obtained via several methods. First, we ran NCBI BLASTX 2.2.16 to compare assembled fire ant sequences with the non-redundant protein database (EMBL release 99). We retained informative gene descriptions of hits with E-value $< 10^{-5}$. Second, Gene Ontology (GO) (Ashburner *et al.* 2000) annotations were inferred using BLASTX as previously described (Wurm *et al.* 2009). Finally, each fire ant sequence was manually curated and assigned a single descriptive category. The manually assigned gene category putatively encapsulates the general function of each sequence and is derived subjectively by examining the SwissProt or Ensembl database entries of the five best BLASTX hits (E-values $< 10^{-5}$), with an emphasis on GO, Interpro, and PANTHER annotations (J. Wang, M. Nicolas and L. Ometto, University of Lausanne, unpublished).

Overrepresentation of manually assigned gene categories and GO categories was determined, respectively, using exact one-sided Fisher tests in R and the Elim test from the topGO Bioconductor package (Alexa *et al.* 2006) limited to categories containing at least 10 fire ant genes.

Comparison with data from other species

Only few complete protein coding gene sequences are available for ants because only partial transcriptome and no genomic sequence data are published. Information about gene orthology between ants and other organisms is practically nonexistent. Nevertheless, we wanted to compare expression

differences identified in this study with similar studies performed in other insects. For this, we modified the Inparanoid ortholog identification pipeline (Remm *et al.* 2001) as follows: BLASTX was replaced with TBLASTX and stringency was reduced so that match areas must span at least 25% of the longer sequence with actual matching segments aligning with at least 10% of the longer sequence. We independently ran this modified Inparanoid pipeline on the assembled fire ant sequences and the complete set of coding sequences of each of the following species: *Drosophila melanogaster* (Flybase release 5.9), *Apis mellifera* (Honey Bee Official Gene Set pre-release 2) and *Anopheles gambiae* (AgamP3.4).

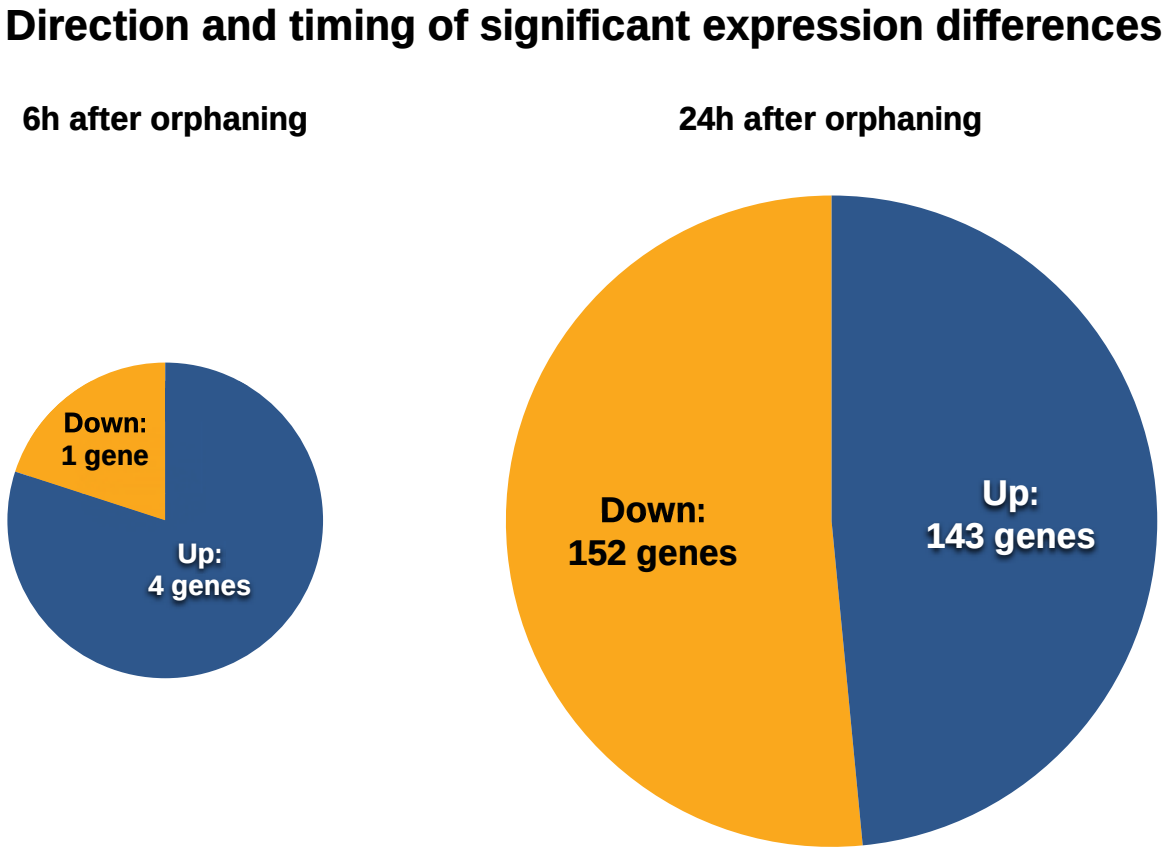
We downloaded lists of significantly over and under-expressed genes from several studies (Lawniczak & Begun 2004; McGraw *et al.* 2004; Rogers *et al.* 2008; Grozinger *et al.* 2007; Kocher *et al.* 2008). The mapping between microarray probes and coding sequences was either provided by the study's authors (for Grozinger *et al.* 2007), obtained by BLASTN of probe sequences to coding sequences (for Kocher *et al.* 2008) or downloaded (for Lawniczak & Begun 2004; McGraw *et al.* 2004; Rogers *et al.* 2008) from BioMart (Haider *et al.* 2009). We were able to compare gene expression levels between our study and each of the other studies for a reduced set of genes that are both putatively orthologous between ants and the species from the other study and present on both the ant microarray and the microarray used in the other study. For the Grozinger *et al.* (2007) study, we compared our results with the list of 221 bee genes that were more highly expressed in queens than in reproductive workers. For the Kocher *et al.* (2008) study, we compared our results with the list of 441 bee genes that were more highly expressed in mated than virgin queens. For the Rogers *et al.* (2008) study, we compared our results with a combined list of 1663 *Anopheles* genes that were either more highly expressed in females 2h, 6h and 24h after mating than in virgins, or more highly expressed 6h than 2h or 24h than 6h after mating. For the McGraw *et al.* (2004) and Lawniczak & Begun (2004) studies, we compared our results with all individual lists of *Drosophila* genes that were differentially expressed in response to different aspects of mating, as well as with a combined list of all mating-response genes they had identified. Exact one-sided Fisher tests quantified the overlap between lists of significant genes from other species and the lists of genes significantly up- and downregulated in orphaned fire ant queens.

5.3 RESULTS

Differential gene expression after orphaning

Four hundred seventy-five of the 18,444 cDNA clones, putatively representing 297 genes, were significantly differentially expressed between the samples of virgin queens collected 0 hours, 6 hours and 24 hours after orphaning (respectively t_{0h}, t_{6h} and t_{24h}). The remaining genes were either expressed similarly before and after orphaning, were highly variable between biological replicates, or yielded signals too weak for reliable assessment of differential expression. Among the 297 significantly differentially expressed genes, four were upregulated within 6 hours of orphaning, while one was downregulated. One hundred forty-four genes were more highly expressed twenty-four hours after orphaning than at $t_{\diagup\!\diagdown}$ or at $t_{\diagup\!\diagup}$ including one of the four genes that was already upregulated after 6 hours, while a total of 152 genes were significantly downregulated after 24 hours (Figure 5.1). One of the genes significantly upregulated after 6 hours was significantly downregulated between 6 and 24 hours. The significant genes are listed in Supporting Tables 1 and 2. These gene expression changes precede or are independent of wing shedding since none of 40 virgin queens collected 6 hours after orphaning and only three of 40 virgin queens collected 24 hours after orphaning had shed their wings.

Figure 5.1: Numbers of genes significantly differentially expressed in young fire ant queens within six hours (left) and 24 hours of orphaning (right).



Gene set enrichment analysis

We bioinformatically annotated the genes that were significantly upregulated or downregulated after orphaning and compared their annotations with the annotations of all genes examined on the microarray by using two different annotation methods. From our in-house annotation categories, two gene categories were overrepresented among upregulated genes. These were *proteasome* (11 observed, 1.2 expected, exact one-sided Fisher test $p = 1 \quad 10^{-7}$) and *protein transport* (10 observed, 1.5 expected, exact one-sided Fisher test $p = 4 \quad 10^{-6}$). No other categories of our in-house annotation were overrepresented among up or downregulated genes. From the BLAST-inferred Gene Ontology categories, several categories were overrepresented among up- and downregulated genes (complete list in Table 5.1). In particular, genes putatively part of the *proteasomal complex* were overrepresented among the upregulated genes (7 observed, 0.7 expected, $p = 0.0003$, topGO Elim test, adjusted for 10% False Discovery Rate (FDR)). Among downregulated genes, those putatively located in *microsomes* and involved in *oxidation reduction* were overrepresented (respectively 6 observed, 0.5 expected, FDR adjusted topGO Elim test $p = 0.0007$, and 14 observed, 3.3 expected, FDR adjusted topGO Elim test$p = 0.0005$). Additionally, genes that putatively have *aromatase activity* were overrepresented among the significantly downregulated genes (5

Table 5.1: Gene Ontology annotations that are significantly enriched among genes that are significantly upregulated or downregulated after orphaning.

Cellular Component

Direction of gene expression change in response to orphaning	GO Identifier	GO Term	Annotated	Significant	Expected	Fold Enrichment	FDR-adjusted p-value
up	GO:0000502	proteasome complex	17	7	0.66	10.6	0.0003
up	GO:0005788	endoplasmic reticulum lumen	10	4	0.39	10.3	0.0314
up	GO:0005829	cytosol	92	10	3.58	2.8	0.0974
up	GO:0005789	endoplasmic reticulum membrane	63	8	2.45	3.3	0.0974
down	GO:0005792	microsome	27	6	0.51	11.8	0.0007
down	GO:0005789	endoplasmic reticulum membrane	63	6	1.2	5.0	0.0544

Biological Process

Direction of gene expression change in response to orphaning	GO Identifier	GO Term	Annotated	Significant	Expected	Fold Enrichment	FDR-adjusted p-value
up	GO:0043161	proteasomal ubiquitin-dependent protein catabolic process	12	5	0.45	11.1	0.0200
down	GO:0055114	oxidation reduction	153	14	3.3	4.2	0.0005

Molecular Function

Direction of gene expression change in response to orphaning	GO Identifier	GO Term	Annotated	Significant	Expected	Fold Enrichment	FDR-adjusted p-value
down	GO:0070330	aromatase activity	12	5	0.29	17.2	0.0014
down	GO:0020037	heme binding	33	6	0.8	7.5	0.0137
down	GO:0016209	antioxidant activity	12	4	0.29	13.8	0.0137

observed, 0.3 expected, FDR adjusted topGO Elim test $p = 0.0014$). In fact, all five of these genes are putative *Cytochrome P450*s.

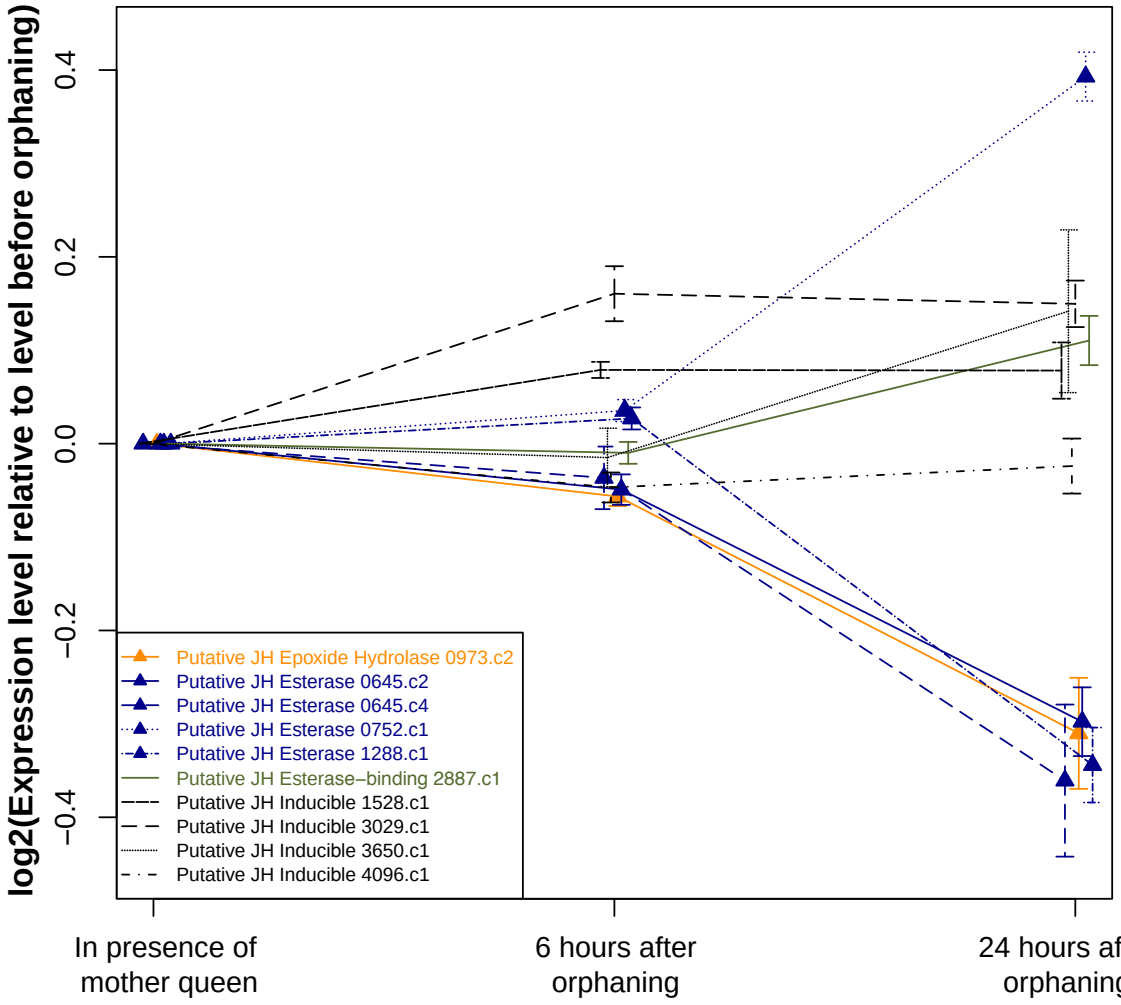
Genes related to Juvenile Hormone metabolism

Among the 297 genes significantly differentially expressed in orphaned compared to non-orphaned queens, five have sequence similarity to genes from other species that are involved in Juvenile Hormone (JH) metabolism or response. In particular, three putative JH esterases were significantly downregulated after orphaning, while one was significantly upregulated. Additionally, a putative JH epoxide hydrolase was significantly downregulated after orphaning. Several putative JH inducible genes as well as a putative JH esterase-binding gene showed non-significant increases in expression level after orphaning (Figure 5.2).

Comparison of fire ant results with data from honey bees

There are two related studies in honey bees allowing comparison with our results. The first one identified genes differentially expressed between brains of honey bee queens and workers (Grozinger *et al.* 2007). We identified a subset of 902 ant-bee orthologs examined in both that study and ours. Genes upregulated in orphaned fire ant queens were enriched for genes upregulated in brains of queen bees relative to brains of reproductive workers (12 observed, 7.5 expected, exact one-sided Fisher test $p = 0.005$). Among the twelve genes that overlap between the groups of significantly upregulated ant and bee genes

Figure 5.2: Expression levels of genes related to Juvenile Hormone (JH) metabolism and response in virgin fire ant queens that are either still in presence of their mother queen or have been orphaned for 6 or 24 hours. Only genes with multiple clones on the microarray are shown. Error bars represent the standard error of the mean expression levels as obtained by independent clones. Genes for which at least one representative clone is significantly differentially expressed after orphaning are indicated by triangles.



(Supporting Table 3), four are part of the manually assigned gene category *proteasome* (0.2 expected, exact one-sided Fisher test $p = 1 * 10^{-4}$).

The other study identified genes differentially expressed between virgin and mated honey bee queens (Kocher *et al.* 2008). Among 2,286 ant-bee orthologs examined in our study as well as the bee study, 13 genes were more highly expressed in response to orphaning in fire ants and in response to mating in honey bee queens (7.7 expected, exact one-sided Fisher test $p = 0.038$, genes listed in Supporting Table 4). Among the thirteen genes that overlap between the two gene lists, four are part of the manually assigned gene category *proteasome* (0.2 expected, exact one-sided Fisher test $p = 1 * 10^{-4}$).

Comparison of fire ant results with data from dipterans

The comparison of our results with those of a study on the effects of mating in female *Anopheles gambiae* mosquitoes for 1,682 orthologs ant-*Anopheles* orthologs (Rogers *et al.* 2008) revealed that genes whose level of expression increased after orphaning in *S. invicta* queens are enriched for genes that are upregulated after mating in *Anopheles* (36 observed, 20.6 expected, exact one-sided Fisher test $p = 8 * 10^{-5}$, genes listed in Supporting Table 5). Six of the thirty-six genes identified in both studies are part of the manually assigned gene category *proteasome* (0.5 expected, exact one-sided Fisher test $p = 3 * 10^{-5}$).

Similar gene expression studies were also performed in the fruitfly *Drosophila melanogaster*. We found no significant overlap between expression changes due to orphaning in fire ant queens and changes due to mating in female *Drosophila* (Lawniczak & Begun 2004; McGraw *et al.* 2004), nor between orphaned fire ant queens and specific aspects of *Drosophila* mating: the mating process itself (without receiving sperm), receiving sperm, or receiving particular accessory proteins normally part of sperm (McGraw *et al.* 2004).

5.4 DISCUSSION

We used microarrays to conduct a genome-wide survey of gene expression in virgin *Solenopsis invicta* fire ant queens over the 24 hours that follow orphaning from their mother queen. We identified five genes that are consistently differentially expressed within six hours of orphaning. These early response genes may be responsible for some of the additional 292 gene expression changes that take place within 24 hours of orphaning. Observing this number of genes showing expression changes is surprising since these changes occur days before the queens shed their wings and begin laying eggs (Vargo & Laurel 1994). However, substantial changes in physiology and metabolism are probably necessary for the transition from a pre-orphaning state where virgin queens are stationary or accumulating lipid reserves to a post-orphaning state in which wing muscles and stored lipids are degraded to develop ovaries and, ultimately, develop eggs. The annotations of the differentially expressed genes indicate that they potentially are involved in many different functions, including signaling reproductive status, reproductive development, proteasomal activity, regulation of transcription and chromatin structure, and protein transport. We discuss each in turn.

Genes potentially involved in signaling of reproductive status

The pheromones that the mother queen uses to signal her presence and fertility are currently unknown. Our study revealed that *Glutathione S-transferase* (GST) is the only gene downregulated in virgin queens 6 hours after orphaning. Furthermore, an additional GST as well as five *Cytochrome P450*s are significantly downregulated in virgin queens within 24 hours of orphaning. Both GSTs and *Cytochrome P450*s are known to be involved in degrading foreign and endogenous compounds (Montellano 2005). We can speculate that these genes may be used by the virgin queens to degrade some of the fertility signals

produced by the mother queen. This could be important if the undesired effect of maternal fertility signals on virgin queens were to trigger reproductive development in the virgins. Alternatively, virgin queens may produce their own fertility signals, and simultaneously degrade them using the GSTs and *Cytochrome P450s*. This would permit young queens to avoid being perceived as a threat to the mother queen when she is present, while simulatenously being able to rapidly increase levels of fertility signals when orphaned.

We also identified three upregulated genes putatively related to olfactory signals, two *chemo-sensory proteins* (CSPs) and one *odorant binding protein* (OBP). These genes could be involved in the perception of potential competing queens. Alternatively, the CSPs and OBP may play the roles of carrier proteins (Gotzek & Ross 2007; Pelosi *et al.* 2005) possibly involved in the production of reproductive status signals. Interestingly, the gene with the highest sequence similarity to the OBP is *Gp-9*, a gene which has been shown to be linked to odor differences between queens (Keller & Ross 1998; Gotzek & Ross 2007) and the selective execution of queens which lack the small *b* allele at this locus in multiple-queen colonies of *S. invicta* (Ross & Keller 1998; Keller & Ross 1998; Ross & Keller 2002; Krieger & Ross 2002). The upregulated OBP could similarly be involved in the production of a qualitative signal by virgin queens.

Genes known to be involved in reproductive development in ants

The level of Juvenile Hormone (JH) has been demonstrated to increase with the onset of reproduction in many female insects including *S. invicta* (Robinson & Vargo 1997; Brent & Vargo 2003). After orphaning, JH synthesis rate increases and JH body content peaks prior to wing shedding (Brent & Vargo 2003). The ectopic application of synthetic JH to virgin queens leads to wing shedding even if the mother queen is present (Vargo & Laurel 1994), whereas applying an inhibitor of JH synthesis represses wing shedding in orphaned virgin queens (Burns *et al.* 2002). The fact that JH level increases after orphaning is consistent with our findings that several genes putatively involved in JH degradation are downregulated after orphaning. Indeed, downregulation of these genes should lead to reduced JH degradation and thus to increased JH levels. Our data also imply that JH degradation genes are highly expressed before orphaning, and thus that JH is already being produced and simultaneously degraded before orphaning. Thus, maintenance of low JH levels in virgin queens prior to orphaning may be due to the simultaneous production and degradation of JH.

Beyond the role of JH, two small-scale studies identified genes associated with reproductive differences in ants. In *S. invicta* queens, participation in a mating flight triggers wing shedding and reproductive development (Tschinkel 2006) and leads to the upregulation of many genes, including seven genes that were identified via sequencing (Tian *et al.* 2004). One of these genes, *Striated Muscle Activator of Rho Signaling* (STARS), was also significantly upregulated in our study 6 and 24 hours after orphaning. Five of the remaining genes, *Vitellogenin-1*, *Vitellogenin-2*, *Yellow-1*, *Yellow-2* and *Abaecin* showed non-significant increases in expression level after orphaning. The expression level of the last gene, *Cytochrome oxidase subunit 2*, remained constant. A study in the black garden ant *Lasius niger* identified seven genes more highly expressed in mature queens than in workers (Gräff *et al.* 2007). Our study revealed that four of these genes were also more highly expressed (but non-significantly so) after orphaning while one showed no change in the level of expression (the last was absent from the *S. invicta* microarray). The overlaps between genes identified in these two studies and in our own suggest that, independently of the species and whether the difference in reproductive status is due to fixed morphological and physiological differences associated with caste or due to a change in adult queens, the same genes may be associated with differences in reproductive status.

Genes that are putatively proteasomal

Genes with similarity to proteasomal genes were highly overrepresented among the genes upregulated after orphaning. Proteasomes are responsible for degrading unneeded proteins. The proteasomal genes could be involved in degrading wing muscle tissue or storage proteins such as hexamerins and vitellogenins that would liberate amino-acids that can be used for reproductive development. Alternatively, the increased proteasomal activity after orphaning may trigger changes in gene expression or cellular proliferation via the respective degradation of transcriptional repressors or specific cyclins. Both possibilities are coherent with the overrepresentation of proteasomal genes among the genes that we identified as being upregulated after orphaning in ant queens and also after mating in bees and mosquitoes. This indicates that since the last common ancestor of ants, bees and mosquitoes (299-359 million years ago, Wiegmann *et al.* 2009), the importance of proteasomal genes during the onset of reproductive development may be evolutionarily conserved. Furthermore, we detected significant downregulation of a gene with similarity to *Cellular Repressor of E1A-stimulated Genes 1* (*CREG1*) after orphaning. *CREG1* has been shown to inhibit growth in human cancer cells and to inhibit apoptosis of human muscle cells (Han *et al.* 2006). The downregulation and degradation of this gene in virgin fire ant queens may similarly induce proliferation of ovarian tissue or the apoptosis of wing muscle cells.

Genes putatively involved in transcriptional changes and chromatin remodeling

Three lines of evidence indicate that widespread transcriptomic and epigenetic changes are taking place after orphaning in virgin fire ant queens. First, the significantly upregulated genes include two putative *RNA polymerase* subunits as well as a putative component of the *Mediator* complex, a coactivator involved in the transcription of nearly all protein-coding genes (Björklund & Gustafsson 2005). Upregulation of these three genes provides evidence for a global increase in gene transcription. Second, a gene containing a Zinc finger transcription factor domain is downregulated, while STARS (*Striated Muscle Activator of Rho Signaling*) and a gene containing a RING finger transcription factor domain are upregulated. These genes may be responsible for many of the gene expression changes we observe. In particular, STARS may induce wing muscle degradation as previously suggested (Tian *et al.* 2004). Alternatively, the differential expression of the three transcriptional activators may indicate that more gene expression changes are imminent. Finally, significant gene expression changes after orphaning include genes that are similar to *Chromobox Homolog protein 1*, also known as *Heterochromatin protein 1* and *Nucleoplasmin-like protein*, also known as *Chromatin decondensation protein*. Both proteins are important for chromatin assembly and disassembly (Lomberk *et al.* 2006; Frehlick *et al.* 2007), and have been shown to play roles in mitosis and meiosis. Some or all of these gene expression changes could be related to the post-orphaning increases in ovarian development and egg production (Vargo & Laurel 1994; Vargo 1999).

Genes putatively involved in protein transport

Genes sharing sequence identity with those involved in protein transport were highly overrepresented among the genes upregulated after orphaning. Proteins need to be shuttled between intracellular compartments for post-translational modifications as well as signal transduction. Protein transportation is also essential for communication between cells via the secretion and uptake of proteins (Lodish *et al.* 2000). The upregulation of putative protein transport genes in orphaned fire ant queens could be involved in changes in neuronal activity (Buckley *et al.* 2000) as a response to orphaning. Alternatively, they may be involved in ovarian development.

CONCLUSION

This study represents the first genome-wide survey of gene expression changes in subordinate animals as they react to the loss of the dominant reproductive. We identified 297 genes differentially expressed within 24 hours of orphaning in virgin *S. invicta* queens. Many of the observed gene expression changes are consistent with previous knowledge about the physiological changes in virgin queens after orphaning, and some genes related to the onset of reproductive development appear to be conserved across species from ants to bees and even mosquitoes. Additionally, we detected several genes possibly required for the perception or production of olfactory signals. These genes may play roles in triggering the onset of reproductive development in virgin queens or in signaling reproductive status to nestmates. Finally, we found evidence for activation of genes putatively involved in muscle degradation and ovarian development. However, much work remains to truly understand the molecular-genetic cascades of events involved in the competition for reproductive dominance between virgin queens. It will be particularly fascinating to understand the evolutionary pressures acting upon different genes involved in this process. A further challenge will be identifying the basis by which workers make decisions regarding which competing queens to execute and which to keep.

ACKNOWLEDGMENTS

We would like to thank Kenneth G Ross and Dietrich Gotzek for help collecting ants, Micah Gardner for help feeding the ants, Christine LaMendola and the Lausanne DNA Array Facility for support with molecular work, Darlene Goldstein, Frédéric Schütz and the Bioconductor mailing list for statistical advice, Rob L Hammond and the Lausanne ant lab for stimulating discussions, and Kenneth G Ross and D. DeWayne Shoemaker for their helpful comments on earlier versions of this manuscript. This work was funded by grants from the Swiss National Science Foundation and the Rectorate of the University of Lausanne.

REFERENCES

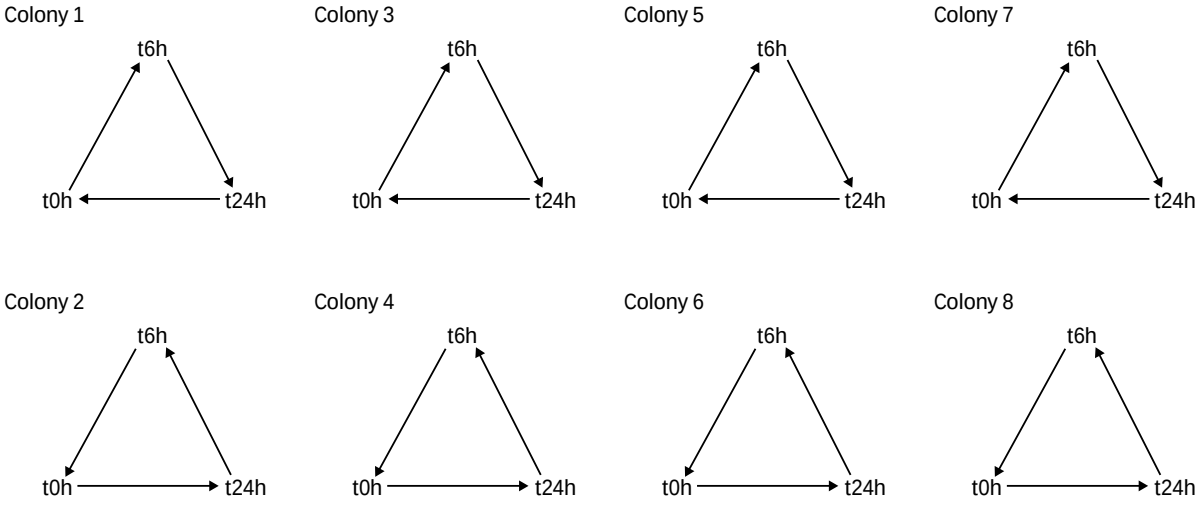
References for this chapter are found on page 113.

SUPPORTING INFORMATION

Supporting information can be downloaded in Excel format from:
http://yannick.poulet.org/miscStuff/MolEcolTimecourseSupplementaryData.zip

Supporting Figure 1:

Graphical representation of microarray hybridizations for queens from colonies 1, 3, 5 and 7 **(a)** and for colonies 2, 4, 6 and 8 **(b)**. Each vertice represents an amplified RNA sample and each edge represents a microarray hybridization. Cy3-labeled samples are at the tails and Cy5-labeled samples are at the heads of arrows.



Supporting Table 1

List of all fire ant genes significantly upregulated for at least one of the following comparisons: 6h vs oh, 24h vs 6h, 24h vs oh. Some genes are significant according to multiple microarray clones.

Clone	"gene" = Contig from March 2009 "E" assembly	Putatively involved in protein transport	Putatively proteasomal	Entry name	Protein names	Organism	evalue	Contig from Oct 2008 Fourmidable assembly	manual_GO1	manual_GO2	manual_annotation
SiJWC11ACR_pcr_2	E.SI.CL.00.cl.0037.Contig1			PIPNA_RABIT	Phosphatidylinositol transfer protein alpha isoform (PtdIns transfer protein alpha) (PtdInsTP) (PI-TP-alpha)	Oryctolagus cuniculus (Rabbit)	3.00E-89	SiJWC11ACR.scf			
SiJWC02ABG_pcr_2	E.SI.CL.00.cl.0038.Contig1			BUD32_DICDI	Probable serine/threonine-protein kinase BUD32 homolog (EC 2.7.11.1)	Dictyostelium discoideum (Slime mold)	4.00E-35	SI.CL.19.cl.1977.SiJWC02ABG.scf	Cell cycle	Cyclins&al positive	Protein kinase
SiJWH09BCY_pcr_1	E.SI.CL.00.cl.0085.Contig1						NA	SI.CL.38.cl.3821.Contig1			
SiJWD04BDG_pcr_1	E.SI.CL.01.cl.0116.Contig1		*	PSMD4_DROME	26S proteasome non-ATPase regulatory subunit 4 (26S proteasome regulatory subunit S5A) (Multiubiquitin chain-binding protein) (54 kDa subunit of mu particle) (p54)	Drosophila melanogaster (Fruit fly)	1.00E-114	SI.CL.20.cl.2007.Contig1	Protein processing	Proteosome	(PSMD4) Proteasome 26S non-ATPase subunit 4
SiJWE07ADT_pcr_1	E.SI.CL.01.cl.0123.Contig1			TUSC3_HUMAN	Tumor suppressor candidate 3 (Protein N33)	Homo sapiens (Human)	1.00E-113	SiJWE07ADT.scf	Protein processing	Chaperon	protein modification
SiJWC11ADQ_pcr_1	E.SI.CL.01.cl.0142.Contig1			RM14_RAT	39S ribosomal protein L14, mitochondrial (L14mt) (MRP-L14) (MRP-L32)	Rattus norvegicus (Rat)	1.00E-30	SI.CL.24.cl.2419.Contig1	Mitochondrion	Translation	Mitochondrial ribosomal protein, L14
SiJWA10AAK_pcr_2	E.SI.CL.01.cl.0151.Contig1						NA	SiJWA10AAK.scf			
SiJWE07AAL_pcr_2	E.SI.CL.01.cl.0175.Contig2						NA	SI.CL.19.cl.1963.SiJWE07AAL.scf			
SiJWG03ACL_pcr_2	E.SI.CL.01.cl.0179.Contig1						NA	SI.CL.7.cl.757.Contig1			
SiJWF04ABE_pcr_2	E.SI.CL.02.cl.0229.Contig2			RS19A_DROME	40S ribosomal protein S19a	Drosophila melanogaster (Fruit fly)	9.00E-49	SiJWF04ABE.tag5_F04_12.scf	Translation	Ribosomal	(rpS19e) Ribosomal protein S19e
SiJWC08BDJ_pcr_1	E.SI.CL.03.cl.0310.Contig1			UB2G2_PONAB	Ubiquitin-conjugating enzyme E2 G2 (EC 6.3.2.19) (Ubiquitin-protein ligase G2) (Ubiquitin carrier protein G2)	Pongo abelii (Sumatran orangutan)	1.00E-74	SI.CL.13.cl.1344.SiJWC08BDJ.scf	Protein processing	Ubiquitine+SUMO	Ubiquitin carrier protein, courtless
SiJWH02BCC_pcr_1	E.SI.CL.03.cl.0379.Contig1			CALR_BOMMO	Calreticulin	Bombyx mori (Silk moth)	5.00E-46	SI.CL.18.cl.1833.Contig1	Uncharacterized	Other	GTP-binding conserved protein with unknown function
SiJWB05BDT_pcr_1	E.SI.CL.03.cl.0379.Contig1			CALR_BOMMO	Calreticulin	Bombyx mori (Silk moth)	5.00E-46	SI.CL.18.cl.1833.SiJWB05BDT2.scf			
SiJWH07ADA_pcr_1	E.SI.CL.03.cl.0379.SiJWH07ADA.scf						NA	SI.CL.18.cl.1833.Contig2			Calreticulin
SiJWE12BDJ_pcr_1	E.SI.CL.04.cl.0437.Contig1			K6PF_DROME	6-phosphofructokinase (Phosphofructokinase) (EC 2.7.1.11) (Phosphohexokinase)	Drosophila melanogaster (Fruit fly)	1.00E-99	SI.CL.39.cl.3991.Contig1	Metabolism	Glycolysis	6-phosphofructokinase
SiJWD02AEA_pcr_1	E.SI.CL.04.cl.0480.Contig1	*		SC61B_PONAB	Protein transport protein Sec61 subunit beta	Pongo abelii (Sumatran orangutan)	2.00E-28	SI.CL.23.cl.2301.Contig1	Trafficking		Transport protein Sec61 beta protein
SiJWE02BAY_pcr_1	E.SI.CL.04.cl.0480.Contig1	*		SC61B_PONAB	Protein transport protein Sec61 subunit beta	Pongo abelii (Sumatran orangutan)	2.00E-28	SI.CL.23.cl.2301.Contig1	Trafficking		Transport protein Sec61 beta protein
SiJWH09CAS_pcr_1	E.SI.CL.04.cl.0480.Contig1	*		SC61B_PONAB	Protein transport protein Sec61 subunit beta	Pongo abelii (Sumatran orangutan)	2.00E-28	SI.CL.23.cl.2301.Contig1	Trafficking		Transport protein Sec61 beta protein
SiJWH11BDK_pcr_1	E.SI.CL.05.cl.0577.Contig1			CBX1_MOUSE	Chromobox protein homolog 1 (Heterochromatin protein 1 homolog beta) (HP1 beta) (Modifier 1 protein) (M31) (Heterochromatin protein p25)	Mus musculus (Mouse)	4.00E-49	SI.CL.0.cl.034.Contig2	Chromatin modification/remodeling	Other	Chromobox homolog 1
SiJWG11BBG_pcr_1	E.SI.CL.06.cl.0679.Contig1			TMCO1_RAT	Transmembrane and coiled-coil domains protein 1 (Meg-2-like protein)	Rattus norvegicus (Rat)	5.00E-61	SI.CL.21.cl.2143.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWG04BBC_pcr_1	E.SI.CL.07.cl.0752.Contig1			ESTF_MYZPE	Esterase FE4 (EC 3.1.1.1) (Carboxylic-ester hydrolase)	Myzus persicae (Peach-potato aphid)	4.00E-80	SI.CL.12.cl.1291.Contig1	Signaling	JH related	Juvenile hormone esterase, putative
SiJWA04ABT_pcr_2	E.SI.CL.08.cl.0833.Contig1			SSRG_RAT	Translocon-associated protein subunit gamma (TRAP-gamma) (Signal sequence receptor subunit gamma) (SSR-gamma)	Rattus norvegicus (Rat)	8.00E-50	SI.CL.15.cl.1562.Contig1	Uncharacterized	Other	conserved unknown protein
SiJWB03ACP_pcr_2	E.SI.CL.08.cl.0833.Contig1			SSRG_RAT	Translocon-associated protein subunit gamma (TRAP-gamma) (Signal sequence receptor subunit gamma) (SSR-gamma)	Rattus norvegicus (Rat)	8.00E-50	SI.CL.15.cl.1562.Contig1	Uncharacterized	Other	conserved unknown protein
SiJWC06BBK_pcr_1	E.SI.CL.08.cl.0833.Contig1			SSRG_RAT	Translocon-associated protein subunit gamma (TRAP-gamma) (Signal sequence receptor subunit gamma) (SSR-gamma)	Rattus norvegicus (Rat)	8.00E-50	SI.CL.15.cl.1562.Contig1	Uncharacterized	Other	conserved unknown protein
SiJWG04AAC_pcr_1	E.SI.CL.08.cl.0833.Contig1			SSRG_RAT	Translocon-associated protein subunit gamma (TRAP-gamma) (Signal sequence receptor subunit gamma) (SSR-gamma)	Rattus norvegicus (Rat)	8.00E-50	SI.CL.15.cl.1562.Contig1	Uncharacterized	Other	conserved unknown protein
SiJWG04AAC_pcr_2	E.SI.CL.08.cl.0833.Contig1			SSRG_RAT	Translocon-associated protein subunit gamma (TRAP-gamma) (Signal sequence receptor subunit gamma) (SSR-gamma)	Rattus norvegicus (Rat)	8.00E-50	SI.CL.15.cl.1562.Contig1	Uncharacterized	Other	conserved unknown protein
SiJWE08ADB_pcr_1	E.SI.CL.09.cl.0936.Contig1		*	PSA5_MOUSE	Proteasome subunit alpha type-5 (EC 3.4.25.1) (Proteasome zeta chain) (Macropain zeta chain) (Multicatalytic endopeptidase complex zeta chain)	Mus musculus (Mouse)	1.00E-107	SI.CL.25.cl.2545.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWD12AAU_pcr_2	E.SI.CL.09.cl.0945.Contig1			TIM10_DROME	Mitochondrial import inner membrane translocase subunit Tim10	Drosophila melanogaster (Fruit fly)	1.00E-26	SI.CL.12.cl.1297.Contig1	Mitochondrion	Other	Mitochondrial import inner membrane translocase subunit Tim10
SiJWC05BBE_pcr_1	E.SI.CL.09.cl.0973.Contig1		*	PSB3_ONCMY	Proteasome subunit beta type-3 (EC 3.4.25.1) (Proteasome theta chain) (Proteasome chain 13) (Proteasome component C10-II)	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	2.00E-84	SI.CL.5.cl.543.Contig1	Protein processing	Proteosome	(Psmb3)Proteasome subunit beta type-3
SiJWD01CAV_pcr_1	E.SI.CL.09.cl.0973.Contig1		*	PSB3_ONCMY	Proteasome subunit beta type-3 (EC 3.4.25.1) (Proteasome theta chain) (Proteasome chain 13) (Proteasome component C10-II)	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	2.00E-84	SI.CL.5.cl.543.Contig1	Protein processing	Proteosome	(Psmb3)Proteasome subunit beta type-3
SiJWD05AAB_pcr_1	E.SI.CL.09.cl.0973.Contig1		*	PSB3_ONCMY	Proteasome subunit beta type-3 (EC 3.4.25.1) (Proteasome theta chain) (Proteasome chain 13) (Proteasome component C10-II)	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	2.00E-84	SI.CL.5.cl.543.Contig1	Protein processing	Proteosome	(Psmb3)Proteasome subunit beta type-3
SiJWD05AAB_pcr_2	E.SI.CL.09.cl.0973.Contig1		*	PSB3_ONCMY	Proteasome subunit beta type-3 (EC 3.4.25.1) (Proteasome theta chain) (Proteasome chain 13) (Proteasome component C10-II)	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	2.00E-84	SI.CL.5.cl.543.Contig1	Protein processing	Proteosome	(Psmb3)Proteasome subunit beta type-3
SiJWE02BBP_pcr_1	E.SI.CL.09.cl.0973.Contig1		*	PSB3_ONCMY	Proteasome subunit beta type-3 (EC 3.4.25.1) (Proteasome theta chain) (Proteasome chain 13) (Proteasome component C10-II)	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	2.00E-84	SI.CL.5.cl.543.Contig1	Protein processing	Proteosome	(Psmb3)Proteasome subunit beta type-3
SiJWG05ACE_pcr_2	E.SI.CL.09.cl.0973.Contig1		*	PSB3_ONCMY	Proteasome subunit beta type-3 (EC 3.4.25.1) (Proteasome theta chain) (Proteasome chain 13) (Proteasome component C10-II)	Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri)	2.00E-84	SI.CL.5.cl.543.Contig1	Protein processing	Proteosome	(Psmb3)Proteasome subunit beta type-3
SiJWC04ACX_pcr_1	E.SI.CL.11.cl.1152.Contig1		*	PSA71_DROME	Proteasome subunit alpha type-7-1 (EC 3.4.25.1) (Proteasome 28 kDa subunit 1) (PROS-Dm28.1)	Drosophila melanogaster (Fruit fly)	3.00E-96	SI.CL.12.cl.1223.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWC07AAV_pcr_2	E.SI.CL.11.cl.1152.Contig1		*	PSA71_DROME	Proteasome subunit alpha type-7-1 (EC 3.4.25.1) (Proteasome 28 kDa subunit 1) (PROS-Dm28.1)	Drosophila melanogaster (Fruit fly)	3.00E-96	SI.CL.12.cl.1223.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWE07ADY_pcr_1	E.SI.CL.11.cl.1152.Contig1		*	PSA71_DROME	Proteasome subunit alpha type-7-1 (EC 3.4.25.1) (Proteasome 28 kDa subunit 1) (PROS-Dm28.1)	Drosophila melanogaster (Fruit fly)	3.00E-96	SI.CL.12.cl.1223.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWE07ADY_pcr_newprimers	E.SI.CL.11.cl.1152.Contig1		*	PSA71_DROME	Proteasome subunit alpha type-7-1 (EC 3.4.25.1) (Proteasome 28 kDa subunit 1) (PROS-Dm28.1)	Drosophila melanogaster (Fruit fly)	3.00E-96	SI.CL.12.cl.1223.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWC09ABR_pcr_2	E.SI.CL.11.cl.1156.Contig1			PRDX5_MOUSE	Peroxiredoxin-5, mitochondrial (EC 1.11.1.15) (Prx-V) (Peroxisomal antioxidant enzyme) (PLP) (Thioredoxin reductase) (Thioredoxin peroxidase PMP20) (Antioxidant enzyme B166) (AOEB166) (Liver tissue 2D-page spot 2D-0014IV)	Mus musculus (Mouse)	1.00E-45	SI.CL.5.cl.579.Contig1			Protein with oxidoreductase activity
SiJWH07AAW_pcr_2	E.SI.CL.11.cl.1156.Contig1			PRDX5_MOUSE	Peroxiredoxin-5, mitochondrial (EC 1.11.1.15) (Prx-V) (Peroxisomal antioxidant enzyme) (PLP) (Thioredoxin reductase) (Thioredoxin peroxidase PMP20) (Antioxidant enzyme B166) (AOEB166) (Liver tissue 2D-page spot 2D-0014IV)	Mus musculus (Mouse)	1.00E-45	SI.CL.5.cl.579.Contig1			Protein with oxidoreductase activity
SiJWB06BCS_pcr_1	E.SI.CL.12.cl.1236.Contig1			PFD3_PONAB	Prefoldin subunit 3 (Von Hippel-Lindau-binding protein 1) (VHL-binding protein 1) (VBP-1)	Pongo abelii (Sumatran orangutan)	1.00E-56	SI.CL.4.cl.407.Contig1	Protein processing	Chaperon	Prefoldin subunit 3
SiJWA05CAI_pcr_1	E.SI.CL.13.cl.1352.Contig1			ADK2_ARATH	Adenosine kinase 2 (AK 2) (EC 2.7.1.20) (Adenosine 5'-phosphotransferase 2)	Arabidopsis thaliana (Mouse-ear cress)	1.00E-111	SI.CL.37.cl.3785.Contig1			Adenosine kinase
SiJWF08ABL_pcr_2	E.SI.CL.13.cl.1352.Contig1			ADK2_ARATH	Adenosine kinase 2 (AK 2) (EC 2.7.1.20) (Adenosine 5'-phosphotransferase 2)	Arabidopsis thaliana (Mouse-ear cress)	1.00E-111	SI.CL.18.cl.1852.Contig1			Adenosine kinase
SiJWH10CAK_pcr_1	E.SI.CL.13.cl.1357.Contig1			CLPT1_DANRE	Cleft lip and palate transmembrane protein 1 homolog	Danio rerio (Zebrafish) (Brachydanio rerio)	1.00E-137	SI.CL.16.cl.1644.Contig3			no hit
SiJWH11AAR_pcr_2	E.SI.CL.14.cl.1426.Contig1	*		PPIB_CHICK	Peptidyl-prolyl cis-trans isomerase B (PPIase) (Rotamase) (EC 5.2.1.8) (Cyclophilin B) (S-cyclophilin) (SCYLP)	Gallus gallus (Chicken)	5.00E-68	SI.CL.12.cl.1215.Contig1	Trafficking		Peptidyl-prolyl cis-trans isomerase
SiJWE07CAF_pcr_1	E.SI.CL.14.cl.1438.Contig1			RN181_BOVIN	E3 ubiquitin-protein ligase RNF181 (RING finger protein 181) (EC 6.3.2.-)	Bos taurus (Bovine)	3.00E-30	SI.CL.0.cl.009.Contig1	Uncharacterized	Zinc finger	RING finger protein
SiJWG11BAQ_pcr_1	E.SI.CL.14.cl.1496.Contig1	*		ARL2_HUMAN	ADP-ribosylation factor-like protein 2	Homo sapiens (Human)	6.00E-67	SI.CL.19.cl.1961.Contig1	Trafficking		ADP-ribosylation factor-like protein
SiJWA07ADO_pcr_1	E.SI.CL.15.cl.1551.Contig1						NA	SiJWA07ADO.scf			
SiJWE11CAL_pcr_1	E.SI.CL.15.cl.1597.Contig1			DENR_DANRE	Density-regulated protein (DRP)	Danio rerio (Zebrafish) (Brachydanio rerio)	4.00E-52				
SiJWB02ADE_pcr_1	E.SI.CL.16.cl.1644.Contig1						NA	SI.CL.21.cl.2114.Contig1			
SiJWA08BDK_pcr_1	E.SI.CL.16.cl.1654.Contig1			SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWB02ADD_pcr_1	E.SI.CL.16.cl.1654.Contig1			SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWC01ADU_pcr_1	E.SI.CL.16.cl.1654.Contig1			SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWC12ACI_pcr_2	E.SI.CL.16.cl.1654.Contig1			SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)

SiJWD06ACL_pcr_2	E.SI.CL.16.cl.1654.Contig1		SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWG01ABQ_pcr_2	E.SI.CL.16.cl.1654.Contig1		SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWG02CAL_pcr_1	E.SI.CL.16.cl.1654.Contig1		SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWG11CAP_pcr_1	E.SI.CL.16.cl.1654.Contig1		SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWH10CAJ_pcr_1	E.SI.CL.16.cl.1654.Contig1		SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)	8.00E-43	SI.CL.2.cl.271.Contig1	Signaling	Other	protein bindng (Translocon-associated protein)
SiJWD09BBX_pcr_1	E.SI.CL.17.cl.1732.Contig1	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7 I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)	3.00E-62	SI.CL.18.cl.1803.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWD10ABW_pcr_2	E.SI.CL.17.cl.1732.Contig1	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7 I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)	3.00E-62	SI.CL.18.cl.1803.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWE01BDT_pcr_1	E.SI.CL.17.cl.1732.Contig1	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7 I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)	3.00E-62	SI.CL.18.cl.1803.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWF10AAA_pcr_1	E.SI.CL.17.cl.1732.Contig1	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7 I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)	3.00E-62	SI.CL.18.cl.1803.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWF10AAA_pcr_2	E.SI.CL.17.cl.1732.Contig1	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7 I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)	3.00E-62	SI.CL.18.cl.1803.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWH08AAY_pcr_1	E.SI.CL.17.cl.1732.Contig1	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7 I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)	3.00E-62	SI.CL.18.cl.1803.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWH08AAY_pcr_2	E.SI.CL.17.cl.1732.Contig1	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7 I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)	3.00E-62	SI.CL.18.cl.1803.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWB01ADF_pcr_1	E.SI.CL.17.cl.1789.Contig1		NDKA_DROYA	Nucleoside diphosphate kinase (NDP kinase) (NDK) (EC 2.7.4.6) (Abnormal wing disks protein) (Fragment)	Drosophila yakuba (Fruit fly)	1.00E-66	SI.CL.4.cl.492.Contig1			Nucleoside diphosphate kinase
SiJWG09ABC_pcr_2	E.SI.CL.17.cl.1789.Contig1		NDKA_DROYA	Nucleoside diphosphate kinase (NDP kinase) (NDK) (EC 2.7.4.6) (Abnormal wing disks protein) (Fragment)	Drosophila yakuba (Fruit fly)	1.00E-66	SI.CL.4.cl.492.Contig1			Nucleoside diphosphate kinase
SiJWF08ADP_pcr_1	E.SI.CL.17.cl.1789.Contig2		NDK_NEUCR	Nucleoside diphosphate kinase (NDP kinase) (NDK) (EC 2.7.4.6)	Neurospora crassa	1.00E-47	SI.CL.4.cl.492.Contig1			Nucleoside diphosphate kinase
SiJWG07BAA_pcr_2	E.SI.CL.17.cl.1789.Contig2		NDK_NEUCR	Nucleoside diphosphate kinase (NDP kinase) (NDK) (EC 2.7.4.6)	Neurospora crassa	1.00E-47	SI.CL.4.cl.492.Contig1			Nucleoside diphosphate kinase
SiJWH12AAX_pcr_1	E.SI.CL.17.cl.1789.Contig2		NDK_NEUCR	Nucleoside diphosphate kinase (NDP kinase) (NDK) (EC 2.7.4.6)	Neurospora crassa	1.00E-47	SI.CL.4.cl.492.Contig1			Nucleoside diphosphate kinase
SiJWE09CAG_pcr_2	E.SI.CL.19.cl.1988.Contig1	*	PSB1_DROME	Proteasome subunit beta type-1 (EC 3.4.25.1) (Proteasome 26 kDa subunit)	Drosophila melanogaster (Fruit fly)	2.00E-71	SI.CL.6.cl.680.Contig1	Protein processing	Proteosome	Proteasome subunit beta type
SiJWA03ABI_pcr_2	E.SI.CL.21.cl.2116.Contig1		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	8.00E-07	SI.CL.12.cl.1214.Contig2	Chemosensory Proteins	Other	Chemosensory protein
SiJWA07BBH_pcr_1	E.SI.CL.21.cl.2116.Contig1		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	8.00E-07	SI.CL.12.cl.1214.Contig2	Chemosensory Proteins	Other	Chemosensory protein
SiJWC06BCP_pcr_1	E.SI.CL.21.cl.2116.Contig1		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	8.00E-07	SI.CL.12.cl.1214.Contig2	Chemosensory Proteins	Other	Chemosensory protein
SiJWG05BAP_pcr_1	E.SI.CL.21.cl.2116.Contig1		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	8.00E-07	SI.CL.12.cl.1214.Contig2	Chemosensory Proteins	Other	Chemosensory protein
SiJWH05ADQ_pcr_1	E.SI.CL.21.cl.2116.Contig1		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	8.00E-07	SI.CL.12.cl.1214.Contig2	Chemosensory Proteins	Other	Chemosensory protein
SiJWB02BDH_pcr_1	E.SI.CL.21.cl.2116.Contig2		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	5.00E-15	SI.CL.12.cl.1214.Contig1	Chemosensory Proteins	Other	Chemosensory protein
SiJWD11BAP_pcr_1	E.SI.CL.21.cl.2116.Contig2		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	5.00E-15	SI.CL.12.cl.1214.Contig1	Chemosensory Proteins	Other	Chemosensory protein
SiJWF11BCE_pcr_1	E.SI.CL.21.cl.2116.Contig2		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	5.00E-15	SI.CL.12.cl.1214.Contig1	Chemosensory Proteins	Other	Chemosensory protein
SiJWG04AAQ_pcr_2	E.SI.CL.21.cl.2116.Contig2		PEB3_DROME	Ejaculatory bulb-specific protein 3 (Ejaculatory bulb-specific protein III) (PEB-melII) (PEBmelIII)	Drosophila melanogaster (Fruit fly)	5.00E-15	SI.CL.12.cl.1214.Contig1	Chemosensory Proteins	Other	Chemosensory protein
SiJWH06BAP_pcr_1	E.SI.CL.22.cl.2202.Contig1		HARB1_RAT	Putative nuclease HARBI1 (EC 3.1.-.-) (Harbinger transposase-derived nuclease)	Rattus norvegicus (Rat)	2.00E-08	SI.CL.28.cl.2816.Contig1	Uncharacterized	Other	uncharacterized protein
SiJWF03AAE_pcr_2	E.SI.CL.22.cl.2291.Contig1		B6E9V3_SOLIN	Odorant-binding protein	Solenopsis invicta (Red imported fire ant) (Solenopsis wagneri)	6.00E-51	SI.CL.0.cl.003.Contig1			
SiJWF02ACV_pcr_2	E.SI.CL.23.cl.2388.Contig1		DCXR_MESAU	L-xylulose reductase (XR) (EC 1.1.1.10) (Dicarbonyl/L-xylulose reductase) (Sperm antigen P26h)	Mesocricetus auratus (Golden hamster)	1.00E-76	SI.CL.17.cl.1746.Contig1			Dicarbonyl/L-xylulose reductase
SiJWH03BBN_pcr_1	E.SI.CL.23.cl.2388.Contig1		DCXR_MESAU	L-xylulose reductase (XR) (EC 1.1.1.10) (Dicarbonyl/L-xylulose reductase) (Sperm antigen P26h)	Mesocricetus auratus (Golden hamster)	1.00E-76	SI.CL.17.cl.1746.Contig1			Dicarbonyl/L-xylulose reductase
SiJWB06ADJ_pcr_1	E.SI.CL.24.cl.2463.Contig1					NA	SiJWB06ADJ.scf			
SiJWA11BBY_pcr_1	E.SI.CL.25.cl.2558.Contig1		B3GLT_HUMAN	Beta-1,3-glucosyltransferase (Beta3Glc-T) (EC 2.4.1.-) (Beta-3-glycosyltransferase-like)	Homo sapiens (Human)	3.00E-17	SiJWA11BBY.scf	Uncharacterized	Other	conserved uncharacterized protein
SiJWB03AAQ_pcr_2	E.SI.CL.26.cl.2601.Contig1		CUD1_SCHGR	Endocuticle structural glycoprotein SgAbd-1	Schistocerca gregaria (Desert locust)	5.00E-17	SI.CL.5.cl.561.SiJWB03AAQ.scf	Cytoskeleton/Matrix		Structural cuticle protein, probable
SiJWC08ABC_pcr_2	E.SI.CL.26.cl.2601.Contig1		CUD1_SCHGR	Endocuticle structural glycoprotein SgAbd-2	Schistocerca gregaria (Desert locust)	5.00E-17	SI.CL.5.cl.561.SiJWC08ABC.scf			no hit
SiJWG02BDJ_pcr_1	E.SI.CL.26.cl.2601.Contig1		CUD1_SCHGR	Endocuticle structural glycoprotein SgAbd-3	Schistocerca gregaria (Desert locust)	5.00E-17	SI.CL.5.cl.561.SiJWG02BDJ.scf			no hit
SiJWH06ADT_pcr_1	E.SI.CL.26.cl.2603.Contig1					NA	SiJWH06ADT.scf			
SiJWE10BDU_pcr_1	E.SI.CL.26.cl.2657.Contig1					NA	SI.CL.3.cl.346.Contig1			
SiJWB07ACA_pcr_2	E.SI.CL.28.cl.2809.Contig1		ABRA_PIG	Actin-binding Rho-activating protein (Striated muscle activator of Rho-dependent signaling) (STARS)	Sus scrofa (Pig)	3.00E-20	SI.CL.3.cl.349.Contig1	Signaling	Other	Striated muscle activator of Rho-dependent signaling
SiJWG10BAO_pcr_1	E.SI.CL.28.cl.2809.Contig1		ABRA_PIG	Actin-binding Rho-activating protein (Striated muscle activator of Rho-dependent signaling) (STARS)	Sus scrofa (Pig)	3.00E-20	SI.CL.3.cl.349.Contig1	Signaling	Other	Striated muscle activator of Rho-dependent signaling
SiJWF12ACQ_pcr_2	E.SI.CL.28.cl.2865.Contig1	*	SC61A_HALRO	Protein transport protein Sec61 subunit alpha	Halocynthia roretzi (Sea squirt)	1.00E-43	SI.CL.8.cl.894.Contig1	Trafficking		
SiJWE03BBR_pcr_1	E.SI.CL.28.cl.2865.Contig2					NA	SI.CL.8.cl.894.Contig1			
SiJWE01ADJ_pcr_1	E.SI.CL.28.cl.2866.Contig1		EBP2_DROME	Probable rRNA-processing protein EBP2 homolog	Drosophila melanogaster (Fruit fly)	1.00E-50	SiJWE01ADJ.scf	Translation	Ribosomal	Probable rRNA-processing protein EBP2
SiJWC03ADQ_pcr_1	E.SI.CL.29.cl.2944.Contig1		ARMET_DROME	ARMET-like protein	Drosophila melanogaster (Fruit fly)	3.00E-62	SI.CL.12.cl.1288.Contig1			
SiJWE12ABY_pcr_2	E.SI.CL.29.cl.2944.Contig1		ARMET_DROME	ARMET-like protein	Drosophila melanogaster (Fruit fly)	3.00E-62	SI.CL.12.cl.1288.Contig1			
SiJWH03ADL_pcr_1	E.SI.CL.30.cl.3003.Contig3		TBB1_MANSE	Tubulin beta-1 chain (Beta-1-tubulin)	Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm)	0	SI.CL.9.cl.919.Contig1	Cytoskeleton/Matrix		Beta tubulin
SiJWF10AEA_pcr_1	E.SI.CL.30.cl.3008.Contig2		RT10_DROME	28S ribosomal protein S10, mitochondrial (S10mt) (MRP-S10)	Drosophila melanogaster (Fruit fly)	5.00E-07	SI.CL.23.cl.2310.Contig1	Mitochondrion	Translation	(mRpS10) Mitochondrial ribosomal protein S10
SiJWG06BCM_pcr_1	E.SI.CL.30.cl.3008.Contig2		RT10_DROME	28S ribosomal protein S10, mitochondrial (S10mt) (MRP-S10)	Drosophila melanogaster (Fruit fly)	5.00E-07	SI.CL.23.cl.2310.Contig1	Mitochondrion	Translation	(mRpS10) Mitochondrial ribosomal protein S10
SiJWD05BAK_pcr_2	E.SI.CL.30.cl.3078.Contig1		GPT_MOUSE	UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase (EC 2.7.8.15) (N-acetylglucosamine-1-phosphate transferase) (GlcNAc-1-P transferase) (G1PT) (GPT)	Mus musculus (Mouse)	4.00E-30	SiJWD05BAK.scf			Glucosaminephosphotransferase, probable
SiJWD06BBQ_pcr_1	E.SI.CL.31.cl.3144.Contig1	*	PSA6_RAT	Proteasome subunit alpha type-6 (EC 3.4.25.1) (Proteasome iota chain) (Macropain iota chain) (Multicatalytic endopeptidase complex iota chain)	Rattus norvegicus (Rat)	1.00E-107	SI.CL.6.cl.663.Contig2	Protein processing	Proteosome	(Psma6) Proteasome subunit alpha type-6
SiJWD09ACD_pcr_2	E.SI.CL.31.cl.3144.Contig1	*	PSA6_RAT	Proteasome subunit alpha type-6 (EC 3.4.25.1) (Proteasome iota chain) (Macropain iota chain) (Multicatalytic endopeptidase complex iota chain)	Rattus norvegicus (Rat)	1.00E-107	SI.CL.6.cl.663.Contig2	Protein processing	Proteosome	(Psma6) Proteasome subunit alpha type-6

SiJWF12CAF_pcr_1	E.SI.CL.31.cl.3144.Contig1	*	PSA6_RAT	Proteasome subunit alpha type-6 (EC 3.4.25.1) (Proteasome iota chain) (Macropain iota chain) (Multicatalytic endopeptidase complex iota chain)	Rattus norvegicus (Rat)	1.00E-107	SI.CL.6.cl.663.Contig2	Protein processing	Proteosome	(Psma6) Proteasome subunit alpha type-6
SiJWH02BDD_pcr_1	E.SI.CL.31.cl.3144.Contig1	*	PSA6_RAT	Proteasome subunit alpha type-6 (EC 3.4.25.1) (Proteasome iota chain) (Macropain iota chain) (Multicatalytic endopeptidase complex iota chain)	Rattus norvegicus (Rat)	1.00E-107	SI.CL.6.cl.663.Contig2	Protein processing	Proteosome	(Psma6) Proteasome subunit alpha type-6
SiJWH10BDA_pcr_1	E.SI.CL.31.cl.3153.Contig1		SERB_BOVIN	Phosphoserine phosphatase (PSPase) (PSP) (EC 3.1.3.3) (O-phosphoserine phosphohydrolase) (L-3-phosphoserine phosphatase)	Bos taurus (Bovine)	1.00E-62	SI.CL.17.cl.1743.Contig1			Phosphoserine phosphatase
SiJWC07BAD_pcr_1	E.SI.CL.32.cl.3212.Contig1		DPH5_HUMAN	Diphthine synthase (EC 2.1.1.98) (Diphthamide biosynthesis methyltransferase)	Homo sapiens (Human)	5.00E-88	SiJWC07BAD.scf	Transcription	Core machinery	Diphthine synthase
SiJWA12AAR_pcr_2	E.SI.CL.32.cl.3230.Contig2		NLP_DROME	Nucleoplasmin-like protein (dNLP) (Chromatin decondensation protein 1)	Drosophila melanogaster (Fruit fly)	3.00E-15	SI.CL.8.cl.806.Contig2	Chromatin modification/remodeling	Decondensation (acethylation) / Elicase	nucleoplasmin, binds to core histones and transfer DNA to them
SiJWF04BDC_pcr_1	E.SI.CL.32.cl.3230.Contig2		NLP_DROME	Nucleoplasmin-like protein (dNLP) (Chromatin decondensation protein 1)	Drosophila melanogaster (Fruit fly)	3.00E-15				
SiJWE03AAX_pcr_1	E.SI.CL.32.cl.3262.Contig1		MRT4_BOVIN	mRNA turnover protein 4 homolog	Bos taurus (Bovine)	3.00E-56	SI.CL.17.cl.1766.Contig1	Translation	Ribosomal	ribosomal protein
SiJWE03AAX_pcr_2	E.SI.CL.32.cl.3262.Contig1		MRT4_BOVIN	mRNA turnover protein 4 homolog	Bos taurus (Bovine)	3.00E-56	SI.CL.17.cl.1766.Contig1	Translation	Ribosomal	ribosomal protein
SiJWE01CAE_pcr_1	E.SI.CL.32.cl.3279.Contig1		EXOS1_MOUSE	3'-5' exoribonuclease CSL4 homolog (EC 3.1.13.-) (Exosome component 1)	Mus musculus (Mouse)	6.00E-45	SiJWE01CAE.scf	Transcript maturation		Exosomal 3'-5' exoribonuclease complex subunit ski4
SiJWB06ADW_pcr_1	E.SI.CL.32.cl.3285.Contig1					1.00E-14	SI.CL.23.cl.2376.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential
SiJWB12AAR_pcr_2	E.SI.CL.32.cl.3285.Contig1					1.00E-14	SI.CL.23.cl.2376.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential
SiJWC12CAD_pcr_1	E.SI.CL.32.cl.3285.Contig1					1.00E-14	SI.CL.23.cl.2376.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential
SiJWG04ADM_pcr_1	E.SI.CL.32.cl.3285.Contig1					1.00E-14	SI.CL.23.cl.2376.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential
SiJWG08BDA_pcr_1	E.SI.CL.32.cl.3285.Contig1					1.00E-14	SI.CL.23.cl.2376.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential
SiJWG11BCE_pcr_1	E.SI.CL.32.cl.3285.Contig1					1.00E-14	SI.CL.23.cl.2376.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential
SiJWA07BDS_pcr_1	E.SI.CL.32.cl.3299.Contig1		RUXE_DROME	Probable small nuclear ribonucleoprotein E (snRNP-E) (Sm protein E) (Sm-E) (SmE)	Drosophila melanogaster (Fruit fly)	3.00E-37	SI.CL.8.cl.850.Contig1	Transcript maturation		small nuclear ribonucleoprotein E
SiJWB12BDC_pcr_1	E.SI.CL.32.cl.3299.Contig1		RUXE_DROME	Probable small nuclear ribonucleoprotein E (snRNP-E) (Sm protein E) (Sm-E) (SmE)	Drosophila melanogaster (Fruit fly)	3.00E-37	SI.CL.8.cl.850.Contig1	Transcript maturation		small nuclear ribonucleoprotein E
SiJWB08ABZ_pcr_2	E.SI.CL.33.cl.3319.Contig1		LIPT_HUMAN	Lipoyltransferase 1, mitochondrial (EC 2.3.1.-) (Lipoate-protein ligase) (Lipoate biosynthesis protein) (Lipoyl ligase)	Homo sapiens (Human)	8.00E-58	SI.CL.5.cl.563.Contig1	Protein processing	Chaperon	lipoyltransferase, probable
SiJWC09ABC_pcr_2	E.SI.CL.33.cl.3319.Contig1		LIPT_HUMAN	Lipoyltransferase 1, mitochondrial (EC 2.3.1.-) (Lipoate-protein ligase) (Lipoate biosynthesis protein) (Lipoyl ligase)	Homo sapiens (Human)	8.00E-58	SI.CL.5.cl.563.Contig1	Protein processing	Chaperon	lipoyltransferase, probable
SiJWG01BAV_pcr_1	E.SI.CL.33.cl.3319.Contig1		LIPT_HUMAN	Lipoyltransferase 1, mitochondrial (EC 2.3.1.-) (Lipoate-protein ligase) (Lipoate biosynthesis protein) (Lipoyl ligase)	Homo sapiens (Human)	8.00E-58	SI.CL.5.cl.563.Contig1	Protein processing	Chaperon	lipoyltransferase, probable
SiJWG01BAQ_pcr_1	E.SI.CL.33.cl.3387.Contig1		RPAB1_PONAB	DNA-directed RNA polymerases I, II, and III subunit RPABC1 (RNA polymerases I, II, and III subunit ABC1) (DNA-directed RNA polymerase II subunit E) (RPB5)	Pongo abelii (Sumatran orangutan)	9.00E-97	SI.CL.2.cl.227.Contig1	Transcription	Core machinery	DNA-directed RNA polymerase II
SiJWB02BCS_pcr_1	E.SI.CL.36.cl.3609.Contig1					NA	SI.CL.33.cl.3324.Contig1			
SiJWC04CAA_pcr_1	E.SI.CL.36.cl.3609.Contig1					NA	SI.CL.33.cl.3324.Contig1			
SiJWC09AAC_pcr_1	E.SI.CL.37.cl.3779.Contig1	*	PSA3_HUMAN	Proteasome subunit alpha type-3 (EC 3.4.25.1) (Proteasome component C8) (Macropain subunit C8) (Multicatalytic endopeptidase complex subunit C8)	Homo sapiens (Human)	2.00E-99	SI.CL.2.cl.290.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type 3 (PSMA3)
SiJWC09AAC_pcr_2	E.SI.CL.37.cl.3779.Contig1	*	PSA3_HUMAN	Proteasome subunit alpha type-3 (EC 3.4.25.1) (Proteasome component C8) (Macropain subunit C8) (Multicatalytic endopeptidase complex subunit C8)	Homo sapiens (Human)	2.00E-99	SI.CL.2.cl.290.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type 3 (PSMA3)
SiJWG06ADW_pcr_1	E.SI.CL.37.cl.3779.Contig1	*	PSA3_HUMAN	Proteasome subunit alpha type-3 (EC 3.4.25.1) (Proteasome component C8) (Macropain subunit C8) (Multicatalytic endopeptidase complex subunit C8)	Homo sapiens (Human)	2.00E-99	SI.CL.2.cl.290.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type 3 (PSMA3)
SiJWH02ACK_pcr_2	E.SI.CL.37.cl.3779.Contig1	*	PSA3_HUMAN	Proteasome subunit alpha type-3 (EC 3.4.25.1) (Proteasome component C8) (Macropain subunit C8) (Multicatalytic endopeptidase complex subunit C8)	Homo sapiens (Human)	2.00E-99	SI.CL.2.cl.290.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type 3 (PSMA3)
SiJWB06ACB_pcr_2	E.SI.CL.38.cl.3821.Contig1		Q17J48_AEDAE	Inosine-uridine preferring nucleoside hydrolase (Fragment)	Aedes aegypti (Yellowfever mosquito) (Culex aegypti)	2.00E-61	SI.CL.4.cl.402.Contig1			Inosine/uridine-preferring nucleoside hydrolase
SiJWB10BCP_pcr_1	E.SI.CL.38.cl.3821.Contig1		Q17J48_AEDAE	Inosine-uridine preferring nucleoside hydrolase (Fragment)	Aedes aegypti (Yellowfever mosquito) (Culex aegypti)	2.00E-61	SI.CL.4.cl.402.Contig1			Inosine/uridine-preferring nucleoside hydrolase
SiJWB11CAE_pcr_1	E.SI.CL.38.cl.3821.Contig1		Q17J48_AEDAE	Inosine-uridine preferring nucleoside hydrolase (Fragment)	Aedes aegypti (Yellowfever mosquito) (Culex aegypti)	2.00E-61	SI.CL.4.cl.402.Contig1			Inosine/uridine-preferring nucleoside hydrolase
SiJWB09ACE_pcr_2	E.SI.CL.38.cl.3891.Contig1					NA	SI.CL.2.cl.207.SiJWB09ACE.SP6.scf			
SiJWA05ABH_pcr_2	E.SI.CL.40.cl.4047.Contig1		B7PMJ8_IXOSC	Putative uncharacterized protein	Ixodes scapularis (Black-legged tick) (Deer tick)	3.00E-07	SiJWA05ABH.scf	Uncharacterized	Other	conserved uncharacterized protein
SiJWB12BEA_pcr_2	E.SI.CL.43.cl.4316.Contig1		B4GEF5_DROPE	GL21826	Drosophila persimilis (Fruit fly)	3.00E-13	SI.CL.23.cl.2327.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWF01ADA_pcr_1	E.SI.CL.43.cl.4316.Contig1		B4GEF5_DROPE	GL21827	Drosophila persimilis (Fruit fly)	3.00E-13	SI.CL.23.cl.2327.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWF10BBC_pcr_1	E.SI.CL.43.cl.4316.Contig1		B4GEF5_DROPE	GL21828	Drosophila persimilis (Fruit fly)	3.00E-13	SI.CL.23.cl.2327.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWB08AAQ_pcr_2	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.SiJWB08AAQ.scf			
SiJWE04ADT_pcr_1	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.SiJWE04ADT.scf			
SiJWE11ADR_pcr_1	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.SiJWE11ADR.scf			
SiJWF04ADW_pcr_1	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.SiJWF04ADW.scf			
SiJWF05AAN_pcr_2	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.SiJWF05AAN.scf			
SiJWF10BCY_pcr_1	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.Contig1			
SiJWF12BBO_pcr_1	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.Contig1			
SiJWH08BAW_pcr_1	E.SI.CL.43.cl.4326.Contig1					NA	SI.CL.1.cl.181.Contig1			
SiJWF03BBJ_pcr_1	E.SI.CL.44.cl.4466.Contig1	*	SEC13_HUMAN	Protein SEC13 homolog (SEC13-related protein) (SEC13-like protein 1)	Homo sapiens (Human)	1.00E-122	SI.CL.14.cl.1413.Contig1	Trafficking		putative SEC13-related protein
SiJWF08ADO_pcr_1	E.SI.CL.45.cl.4520.Contig1					NA	SiJWF08ADO.scf			
SiJWF03BAB_pcr_1	E.SI.CL.45.cl.4597.Contig1		RM20_XENLA	39S ribosomal protein L20, mitochondrial (L20mt) (MRP-L20)	Xenopus laevis (African clawed frog)	9.00E-23	SI.CL.28.cl.2832.Contig1	Mitochondrion	Translation	(mRpL20) Ribosomal protein L20, mitochondrial
SiJWG02CAB_pcr_1	E.SI.CL.45.cl.4597.Contig1		RM20_XENLA	39S ribosomal protein L20, mitochondrial (L20mt) (MRP-L20)	Xenopus laevis (African clawed frog)	9.00E-23	SI.CL.28.cl.2832.Contig1	Mitochondrion	Translation	(mRpL20) Ribosomal protein L20, mitochondrial
SiJWA09BAE_pcr_1	E.SI.CL.47.cl.4726.Contig1		S61G1_GRYOR	Protein transport protein Sec61 subunit gamma	Gryllotalpa orientalis (Oriental mole cricket)	3.00E-15	SI.CL.26.cl.2623.Contig1			Protein transport protein Sec61 subunit gamma
SiJWD01AAN_pcr_2	E.SI.CL.47.cl.4768.Contig1		PDIA6_MOUSE	Protein disulfide-isomerase A6 (EC 5.3.4.1) (Thioredoxin domain-containing protein 7)	Mus musculus (Mouse)	1.00E-144	SI.CL.2.cl.201.Contig1			
SiJWE05ABI_pcr_2	E.SI.CL.47.cl.4768.Contig1		PDIA6_MOUSE	Protein disulfide-isomerase A6 (EC 5.3.4.1) (Thioredoxin domain-containing protein 7)	Mus musculus (Mouse)	1.00E-144	SI.CL.2.cl.201.SiJWE05ABI.scf			
SiJWF05ABQ_pcr_2	E.SI.CL.47.cl.4768.Contig1		PDIA6_MOUSE	Protein disulfide-isomerase A6 (EC 5.3.4.1) (Thioredoxin domain-containing protein 7)	Mus musculus (Mouse)	1.00E-144	SI.CL.2.cl.201.SiJWF05ABQ.scf			
SiJWH05AAB_pcr_1	E.SI.CL.47.cl.4786.Contig1		LACB2_XENLA	Beta-lactamase-like protein 2 (EC 3.-.-.-)	Xenopus laevis (African clawed frog)	7.00E-77	SiJWH05AAB.scf			Beta-lactamase-like protein
SiJWC06BCX_pcr_1	E.SI.CL.48.cl.4838.Contig1		NOP16_HUMAN	Nucleolar protein 16 (HBV pre-S2 trans-regulated protein 3)	Homo sapiens (Human)	6.00E-16	SI.CL.4.cl.429.Contig1	Protein processing	Protease	serine protease, putative
SiJWD12AAT_pcr_2	E.SI.CL.48.cl.4862.Contig1		UFC1_RAT	Ufm1-conjugating enzyme 1 (Ubiquitin-fold modifier-conjugating enzyme 1)	Rattus norvegicus (Rat)	3.00E-80	SI.CL.2.cl.269.Contig1	Protein processing	Chaperon	(UFC1) Ufm1-conjugating enzyme 1
SiJWF04ADK_pcr_1	E.SI.CL.48.cl.4862.Contig1		UFC1_RAT	Ufm1-conjugating enzyme 1 (Ubiquitin-fold modifier-conjugating enzyme 1)	Rattus norvegicus (Rat)	3.00E-80	SI.CL.2.cl.269.Contig1	Protein processing	Chaperon	(UFC1) Ufm1-conjugating enzyme 1
SiJWA06BDC_pcr_1	E.SI.CL.49.cl.4907.Contig1		HEXA_PONAB	Beta-hexosaminidase subunit alpha (EC 3.2.1.52) (N-acetyl-beta-glucosaminidase subunit alpha) (Beta-N-acetylhexosaminidase subunit alpha) (Hexosaminidase subunit A)	Pongo abelii (Sumatran orangutan)	7.00E-14	SI.CL.37.cl.3738.Contig1	Metabolism	Glycolysis	Beta-N-acetylglucosaminidase
SiJWE12BCD_pcr_1	E.SI.CL.49.cl.4982.Contig1		PDI_DROME	Protein disulfide-isomerase (PDI) (dPDI) (EC 5.3.4.1)	Drosophila melanogaster (Fruit fly)	7.00E-88	SI.CL.23.cl.2344.Contig1	Protein processing		possible Protein disulfide isomerase
SiJWH06BAT_pcr_1	E.SI.CL.49.cl.4982.Contig1		PDI_DROME	Protein disulfide-isomerase (PDI) (dPDI) (EC 5.3.4.1)	Drosophila melanogaster (Fruit fly)	7.00E-88	SI.CL.23.cl.2344.Contig1	Protein processing		possible Protein disulfide isomerase

SiJWC04ACU_pcr_2	E.SI.CL.50.cl.5004.Contig1		*	PSA4_HUMAN	Proteasome subunit alpha type-4 (EC 3.4.25.1) (Proteasome component C9) (Macropain subunit C9) (Multicatalytic endopeptidase complex subunit C9) (Proteasome subunit L)	Homo sapiens (Human)	1.00E-112	SI.CL.9.cl.992.SiJWC04ACU.scf	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWD05ADL_pcr_1	E.SI.CL.50.cl.5004.Contig1		*	PSA4_HUMAN	Proteasome subunit alpha type-4 (EC 3.4.25.1) (Proteasome component C9) (Macropain subunit C9) (Multicatalytic endopeptidase complex subunit C9) (Proteasome subunit L)	Homo sapiens (Human)	1.00E-112	SI.CL.9.cl.992.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWF05BCN_pcr_1	E.SI.CL.50.cl.5004.Contig1		*	PSA4_HUMAN	Proteasome subunit alpha type-4 (EC 3.4.25.1) (Proteasome component C9) (Macropain subunit C9) (Multicatalytic endopeptidase complex subunit C9) (Proteasome subunit L)	Homo sapiens (Human)	1.00E-112	SI.CL.9.cl.992.Contig2	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWE09BCF_pcr_1	E.SI.CL.50.cl.5004.Contig2			PSA4_RAT	Proteasome subunit alpha type-4 (EC 3.4.25.1) (Proteasome component C9) (Macropain subunit C9) (Multicatalytic endopeptidase complex subunit C9) (Proteasome subunit L)	Rattus norvegicus (Rat)	1.00E-32	SI.CL.9.cl.992.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type
SiJWE10ADV_pcr_1	E.SI.CL.50.cl.5017.Contig1			CD63_BOVIN	CD63 antigen (CD antigen CD63)	Bos taurus (Bovine)	2.00E-19	SiJWE10ADV.scf	Cytoskeleton/Matrix		Tetraspanin, probable
SiJWF08ADS_pcr_1	E.SI.CL.50.cl.5031.Contig1			A8N8V2_COPC7	Putative uncharacterized protein	Coprinopsis cinerea (strain Okayama-7 / 130 / FGSC 9003) (Inky cap fungus) (Hormographiella aspergillata)	7.00E-06	SI.CL.15.cl.1520.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWH02BCX_pcr_1	E.SI.CL.50.cl.5031.Contig1			A8N8V2_COPC7	Putative uncharacterized protein	Coprinopsis cinerea (strain Okayama-7 / 130 / FGSC 9003) (Inky cap fungus) (Hormographiella aspergillata)	7.00E-06	SI.CL.15.cl.1520.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWB06BCG_pcr_1	E.SI.CL.50.cl.5033.Contig1			CH10_SCHJA	10 kDa heat shock protein, mitochondrial (Hsp10) (10 kDa chaperonin) (CPN10)	Schistosoma japonicum (Blood fluke)	1.00E-29	SI.CL.20.cl.2028.Contig1	Protein processing	Chaperon	heat shock protein 10
SiJWB10BAD_pcr_1	E.SI.CL.50.cl.5057.Contig1			MED9_DROPS	Mediator of RNA polymerase II transcription subunit 9 (Mediator complex subunit 9)	Drosophila pseudoobscura pseudoobscura (Fruit fly)	2.00E-17	SI.CL.7.cl.755.Contig1	Transcription	Core machinery	Mediator of RNA polymerase II transcription subunit 9
SiJWC12BBQ_pcr_1	E.SI.CL.50.cl.5057.Contig1			MED9_DROPS	Mediator of RNA polymerase II transcription subunit 9 (Mediator complex subunit 9)	Drosophila pseudoobscura pseudoobscura (Fruit fly)	2.00E-17	SI.CL.7.cl.755.Contig1	Transcription	Core machinery	Mediator of RNA polymerase II transcription subunit 9
SiJWG11ADQ_pcr_1	E.SI.CL.50.cl.5057.Contig1			MED9_DROPS	Mediator of RNA polymerase II transcription subunit 9 (Mediator complex subunit 9)	Drosophila pseudoobscura pseudoobscura (Fruit fly)	2.00E-17	SI.CL.7.cl.755.Contig1	Transcription	Core machinery	Mediator of RNA polymerase II transcription subunit 9
SiJWB08AAM_pcr_2	E.SI.CL.50.cl.5069.Contig1	*		RAB1A_LYMST	Ras-related protein Rab-1A	Lymnaea stagnalis (Great pond snail)	5.00E-95	SiJWB08AAM.scf	Trafficking		
SiJWG09ADW_pcr_1	E.SI.CL.51.cl.5105.SiJWG09ADW.scf			B0W8P6_CULQU	Putative uncharacterized protein	Culex quinquefasciatus (Southern house mosquito) (Culex pungens)	5.00E-12	SI.CL.13.cl.1370.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWF06ADC_pcr_1	E.SI.CL.51.cl.5107.Contig1			IVBIT_ERIMA	Venom trypsin inhibitor	Eristocophis macmahoni (Leaf-nosed viper)	5.00E-09	SI.CL.4.cl.412.SiJWF06ADC.scf	Protein processing	Protease	protease inhibitor
SiJWD06BCU_pcr_1	E.SI.CL.51.cl.5107.Contig2			IVBIC_VIPAM	Venom basic protease inhibitor 3 (Venom chymotrypsin inhibitor) (cVamChi)	Vipera ammodytes (Sand viper)	1.00E-11	SI.CL.4.cl.412.Contig1	Protein processing	Protease	protease inhibitor
SiJWF04ADV_pcr_1	E.SI.CL.51.cl.5107.Contig2			IVBIC_VIPAM	Venom basic protease inhibitor 3 (Venom chymotrypsin inhibitor) (cVamChi)	Vipera ammodytes (Sand viper)	1.00E-11	SI.CL.4.cl.412.Contig1	Protein processing	Protease	protease inhibitor
SiJWG02AAK_pcr_2	E.SI.CL.51.cl.5107.Contig2			IVBIC_VIPAM	Venom basic protease inhibitor 3 (Venom chymotrypsin inhibitor) (cVamChi)	Vipera ammodytes (Sand viper)	1.00E-11	SI.CL.4.cl.412.Contig1	Protein processing	Protease	protease inhibitor
SiJWG06AAF_pcr_2	E.SI.CL.51.cl.5107.Contig2			IVBIC_VIPAM	Venom basic protease inhibitor 3 (Venom chymotrypsin inhibitor) (cVamChi)	Vipera ammodytes (Sand viper)	1.00E-11	SI.CL.4.cl.412.Contig1	Protein processing	Protease	protease inhibitor
SiJWG09BBT_pcr_1	E.SI.CL.51.cl.5107.Contig2			IVBIC_VIPAM	Venom basic protease inhibitor 3 (Venom chymotrypsin inhibitor) (cVamChi)	Vipera ammodytes (Sand viper)	1.00E-11	SI.CL.4.cl.412.Contig1	Protein processing	Protease	protease inhibitor
SiJWH01BAF_pcr_1	E.SI.CL.51.cl.5107.Contig2			IVBIC_VIPAM	Venom basic protease inhibitor 3 (Venom chymotrypsin inhibitor) (cVamChi)	Vipera ammodytes (Sand viper)	1.00E-11	SI.CL.4.cl.412.Contig1	Protein processing	Protease	protease inhibitor
SiJWC08BCU_pcr_1	E.SI.CL.51.cl.5109.Contig1			SPCS3_DROME	Signal peptidase complex subunit 3 (EC 3.4.-.-) (Microsomal signal peptidase 22 kDa subunit) (SPase 22 kDa subunit) (SPC22)	Drosophila melanogaster (Fruit fly)	3.00E-78	SI.CL.3.cl.381.Contig1			
SiJWF06ACB_pcr_2	E.SI.CL.51.cl.5109.Contig1			SPCS3_DROME	Signal peptidase complex subunit 3 (EC 3.4.-.-) (Microsomal signal peptidase 22 kDa subunit) (SPase 22 kDa subunit) (SPC22)	Drosophila melanogaster (Fruit fly)	3.00E-78	SI.CL.3.cl.381.Contig1			
SiJWE03CAT_pcr_1	E.SI.CL.51.cl.5198.Contig1			SYFA_DROME	Probable phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20) (Phenylalanine--tRNA ligase alpha chain) (PheRS)	Drosophila melanogaster (Fruit fly)	3.00E-23	SiJWE03CAT2.scf			
SiJWH09ABL_pcr_2	E.SI.CL.52.cl.5214.Contig1	*		VPS45_RAT	Vacuolar protein sorting-associated protein 45 (rvps45)	Rattus norvegicus (Rat)	1.00E-134	SI.CL.18.cl.1858.Contig1	Trafficking		
SiJWA04BBG_pcr_1	E.SI.CL.52.cl.5241.Contig1			HS90A_PIG	Heat shock protein HSP 90-alpha (HSP 86)	Sus scrofa (Pig)	1.00E-153	SI.CL.0.cl.078.Contig1	Protein processing	Chaperon	(hsp90) 90 kDa heat shock protein
SiJWF07ACF_pcr_2	E.SI.CL.52.cl.5241.Contig1			HS90A_PIG	Heat shock protein HSP 90-alpha (HSP 86)	Sus scrofa (Pig)	1.00E-153	SI.CL.0.cl.078.Contig1	Protein processing	Chaperon	(hsp90) 90 kDa heat shock protein
SiJWH02ACD_pcr_2	E.SI.CL.52.cl.5241.Contig1			HS90A_PIG	Heat shock protein HSP 90-alpha (HSP 86)	Sus scrofa (Pig)	1.00E-153	SI.CL.0.cl.078.Contig1	Protein processing	Chaperon	(hsp90) 90 kDa heat shock protein
SiJWH06BAD_pcr_1	E.SI.CL.52.cl.5264.Contig1						NA	SiJWH06BAD.scf			
SiJWD04ACS_pcr_1	E.SI.CL.53.cl.5305.Contig1			RPP29_RAT	Ribonuclease P protein subunit p29 (EC 3.1.26.5)	Rattus norvegicus (Rat)	3.00E-15	SI.CL.9.cl.957.Contig1			Ribonuclease P/MRP, p29 subunit
SiJWD04ACS_pcr_2	E.SI.CL.53.cl.5305.Contig1			RPP29_RAT	Ribonuclease P protein subunit p29 (EC 3.1.26.5)	Rattus norvegicus (Rat)	3.00E-15	SI.CL.9.cl.957.Contig1			Ribonuclease P/MRP, p29 subunit
SiJWF12CAA_pcr_1	E.SI.CL.53.cl.5305.Contig1			RPP29_RAT	Ribonuclease P protein subunit p29 (EC 3.1.26.5)	Rattus norvegicus (Rat)	3.00E-15	SI.CL.9.cl.957.Contig1			Ribonuclease P/MRP, p29 subunit
SiJWH08BBC_pcr_1	E.SI.CL.53.cl.5362.Contig1	*		CLCA_RAT	Clathrin light chain A (Lca)	Rattus norvegicus (Rat)	3.00E-20	SI.CL.2.cl.272.Contig1	Trafficking		Clathrin light chain
SiJWB09BDR_pcr_1	E.SI.CL.53.cl.5384.Contig1	*		ARF2_DROME	ADP-ribosylation factor 2 (dARF II)	Drosophila melanogaster (Fruit fly)	4.00E-91	SI.CL.12.cl.1235.Contig1	Trafficking		ADP-ribosylation factor 2
SiJWG03AAV_pcr_2	E.SI.CL.53.cl.5384.Contig1	*		ARF2_DROME	ADP-ribosylation factor 2 (dARF II)	Drosophila melanogaster (Fruit fly)	4.00E-91	SI.CL.12.cl.1235.Contig1	Trafficking		ADP-ribosylation factor 2
SiJWA11ADF_pcr_1	E.SI.CL.53.cl.5384.Contig1	*		ARF2_DROME	ADP-ribosylation factor 2 (dARF II)	Drosophila melanogaster (Fruit fly)	4.00E-91	SI.CL.21.cl.2136.Contig1	Trafficking		ADP-ribosylation factor 2
SiJWG12BDL_pcr_1	E.SI.CL.53.cl.5384.Contig1	*		ARF2_DROME	ADP-ribosylation factor 2 (dARF II)	Drosophila melanogaster (Fruit fly)	4.00E-91	SI.CL.21.cl.2136.Contig1	Trafficking		ADP-ribosylation factor 2
SiJWC05BCX_pcr_1	E.SI.CL.53.cl.5392.Contig1			HMDH_BLAGE	3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMG-CoA reductase) (EC 1.1.1.34)	Blattella germanica (German cockroach)	5.00E-53	SI.CL.41.cl.4153.Contig1			
SiJWF09BDU_pcr_1	E.SI.CL.54.cl.5465.Contig1			SERP2_MOUSE	Stress-associated endoplasmic reticulum protein 2 (Ribosome-associated membrane protein RAMP4-2)	Mus musculus (Mouse)	3.00E-18				
SiJWE03AAP_pcr_2	E.SI.CL.55.cl.5504.SiJWE03AAP.scf						NA	SiJWE03AAP.scf			
SiJWB12BDY_pcr_1	E.SI.CL.55.cl.5529.Contig1			TIM22_DROME	Mitochondrial import inner membrane translocase subunit Tim22	Drosophila melanogaster (Fruit fly)	3.00E-56	SI.CL.12.cl.1222.Contig1	Mitochondrion	Other	Mitochondrial import inner membrane translocase subunit Tim22
SiJWC05AAV_pcr_2	E.SI.CL.55.cl.5529.Contig1			TIM22_DROME	Mitochondrial import inner membrane translocase subunit Tim23	Drosophila melanogaster (Fruit fly)	3.00E-56	SI.CL.12.cl.1222.Contig1	Mitochondrion	Other	Mitochondrial import inner membrane translocase subunit Tim22
SiJWH01CAQ_pcr_1	E.SI.CL.55.cl.5529.Contig1			TIM22_DROME	Mitochondrial import inner membrane translocase subunit Tim24	Drosophila melanogaster (Fruit fly)	3.00E-56				
SiJWA04BDQ_pcr_1	E.SI.CL.56.cl.5601.Contig1			MTAP_DROME	Putative S-methyl-5'-thioadenosine phosphorylase (EC 2.4.2.28) (5'-methylthioadenosine phosphorylase) (MTA phosphorylase) (MTAPase)	Drosophila melanogaster (Fruit fly)	4.00E-66	SI.CL.31.cl.3193.Contig1			S-methyl-5-thioadenosine phosphorylase
SiJWF10BDH_pcr_1	E.SI.CL.56.cl.5601.Contig1			MTAP_DROME	Putative S-methyl-5'-thioadenosine phosphorylase (EC 2.4.2.28) (5'-methylthioadenosine phosphorylase) (MTA phosphorylase) (MTAPase)	Drosophila melanogaster (Fruit fly)	4.00E-66	SI.CL.31.cl.3193.Contig1			S-methyl-5-thioadenosine phosphorylase
SiJWB02BDC_pcr_1	E.SI.CL.56.cl.5644.Contig1						NA	SI.CL.33.cl.3308.Contig1			
SiJWA04ACY_pcr_1	E.SI.CL.57.cl.5727.Contig1			RPE_HUMAN	Ribulose-phosphate 3-epimerase (EC 5.1.3.1) (Ribulose-5-phosphate-3-epimerase)	Homo sapiens (Human)	2.00E-66	SiJWA04ACY.scf	Metabolism	Glycolysis	Ribulose-5-phosphate-3-epimerase
SiJWH07CAA_pcr_1	E.SI.CL.58.cl.5866.Contig1			SPCS2_XENTR	Probable signal peptidase complex subunit 2 (EC 3.4.-.-) (Microsomal signal peptidase 25 kDa subunit) (SPase 25 kDa subunit)	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)	2.00E-44	SI.CL.21.cl.2130.Contig1	Protein processing	Protease	Signal peptidase complex subunit 2
SiJWG11AAY_pcr_1	E.SI.CL.58.cl.5889.Contig1			COX11_MOUSE	Cytochrome c oxidase assembly protein COX11, mitochondrial	Mus musculus (Mouse)	5.00E-58	SI.CL.18.cl.1801.Contig1	Mitochondrion	ATP synthesis	Cytochrome c oxidase assembly protein Cox11
SiJWC06ABC_pcr_2	E.SI.CL.59.cl.5912.Contig1						NA	SiJWC06ABC.scf			
SiJWC08CAG_pcr_1	E.SI.CL.59.cl.5920.Contig1						NA	SiJWC08CAG.scf			
SiJWE09ACV_pcr_2	E.SI.CL.60.cl.6036.Contig1	*		PRAF1_HUMAN	Prenylated Rab acceptor protein 1 (PRA1 family protein 1)	Homo sapiens (Human)	6.00E-23	SiJWE09ACV.scf	Trafficking		possible Prenylated RAB acceptor 1
SiJWC05BDS_pcr_1	E.SI.CL.60.cl.6096.Contig1			HYOU1_CRIGR	Hypoxia up-regulated protein 1 (150 kDa oxygen-regulated protein) (Orp150) (170 kDa glucose-regulated protein)	Cricetulus griseus (Chinese hamster)	4.00E-23	SI.CL.34.cl.3467.Contig1	Protein processing	Chaperon	hsp70 domain, probable
SiJWG10ADR_pcr_1	E.SI.CL.61.cl.6155.Contig1			GLRX3_XENTR	Glutaredoxin-3 (Thioredoxin-like protein 2)	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)	2.00E-58	SiJWG10ADR.scf			Glutaredoxin
SiJWG08ADG_pcr_1	E.SI.CL.62.cl.6225.Contig1		*	PSB5_RAT	Proteasome subunit beta type-5 (EC 3.4.25.1) (Proteasome epsilon chain) (Macropain epsilon chain) (Multicatalytic endopeptidase complex epsilon chain) (Proteasome subunit X) (Proteasome chain 6)	Rattus norvegicus (Rat)	5.00E-79	SiJWG08ADG.scf	Protein processing	Proteosome	Proteasome subunit beta
SiJWA05BDW_pcr_2	E.SI.CL.62.cl.6275.Contig1			TIM9_DROME	Mitochondrial import inner membrane translocase subunit Tim9	Drosophila melanogaster (Fruit fly)	4.00E-26	SI.CL.30.cl.3064.Contig1	Mitochondrion	Other	Mitochondrial import inner membrane translocase subunit Tim9
SiJWG02BBC_pcr_1	E.SI.CL.62.cl.6275.Contig1			TIM9_DROME	Mitochondrial import inner membrane translocase subunit Tim10	Drosophila melanogaster (Fruit fly)	4.00E-26	SI.CL.30.cl.3064.Contig1	Mitochondrion	Other	Mitochondrial import inner membrane translocase subunit Tim9
SiJWE03AA1_pcr_1	E.SI.CL.62.cl.6284.Contig1			RPAB2_DROME	DNA-directed RNA polymerases I, II, and III subunit RPABC2 (RNA polymerases I, II, and III subunit ABC2) (RPB6)	Drosophila melanogaster (Fruit fly)	9.00E-39	SI.CL.7.cl.702.Contig1	Transcription	Core machinery	DNA-directed RNA polymerases I, II, and III subunit RPABC2
SiJWA12BBC_pcr_1	E.SI.CL.63.cl.6329.Contig1						NA	SiJWA12BBC2.scf			

SiJWC08BAW_pcr_1	E.SI.CL.63.cl.6329.Contig1						NA	SI.CL.28.cl.2882.Contig1			
SiJWB02BAY_pcr_1	E.SI.CL.63.cl.6332.Contig1		DERL1_PONAB	Derlin-1 (Der1-like protein 1) (Degradation in endoplasmic reticulum protein 1)	Pongo abelii (Sumatran orangutan)	4.00E-82	SI.CL.4.cl.460.Contig1	Protein processing	Proteosome	Degradation in endoplasmic reticulum protein 1	
SiJWD03AAG_pcr_2	E.SI.CL.63.cl.6332.Contig1		DERL1_PONAB	Derlin-1 (Der1-like protein 1) (Degradation in endoplasmic reticulum protein 1)	Pongo abelii (Sumatran orangutan)	4.00E-82	SI.CL.4.cl.460.Contig1	Protein processing	Proteosome	Degradation in endoplasmic reticulum protein 1	
SiJWE06BBR_pcr_1	E.SI.CL.63.cl.6332.Contig1		DERL1_PONAB	Derlin-1 (Der1-like protein 1) (Degradation in endoplasmic reticulum protein 1)	Pongo abelii (Sumatran orangutan)	4.00E-82	SI.CL.4.cl.460.Contig1	Protein processing	Proteosome	Degradation in endoplasmic reticulum protein 1	
SiJWH01AAP_pcr_2	E.SI.CL.63.cl.6332.Contig1		DERL1_PONAB	Derlin-1 (Der1-like protein 1) (Degradation in endoplasmic reticulum protein 1)	Pongo abelii (Sumatran orangutan)	4.00E-82	SI.CL.4.cl.460.Contig1	Protein processing	Proteosome	Degradation in endoplasmic reticulum protein 1	
SiJWC04ACN_pcr_2	E.SI.CL.63.cl.6380.Contig1		RU2A_DROME	Probable U2 small nuclear ribonucleoprotein A' (U2 snRNP-A')	Drosophila melanogaster (Fruit fly)	9.00E-78	SI.CL.8.cl.813.SiJWC04ACN.scf	Transcript maturation		U2 small nuclear ribonucleoprotein A	
SiJWE10AAQ_pcr_2	E.SI.CL.63.cl.6380.Contig1		RU2A_DROME	Probable U2 small nuclear ribonucleoprotein A' (U2 snRNP-A')	Drosophila melanogaster (Fruit fly)	9.00E-78	SI.CL.8.cl.813.SiJWE10AAQ.scf	Transcript maturation		U2 small nuclear ribonucleoprotein A	
SiJWE05ADT_pcr_1	E.SI.CL.64.cl.6455.Contig1		CNPY2_MOUSE	Protein canopy homolog 2 (MIR-interacting saposin-like protein) (Transmembrane protein 4) (Putative secreted protein ZSIG9)	Mus musculus (Mouse)	6.00E-17	SI.CL.2.cl.208.Contig1			putative ER protein	
SiJWE09AAX_pcr_1	E.SI.CL.64.cl.6455.Contig1		CNPY2_MOUSE	Protein canopy homolog 2 (MIR-interacting saposin-like protein) (Transmembrane protein 4) (Putative secreted protein ZSIG9)	Mus musculus (Mouse)	6.00E-17	SI.CL.2.cl.208.Contig1			putative ER protein	
SiJWE09AAX_pcr_2	E.SI.CL.64.cl.6455.Contig1		CNPY2_MOUSE	Protein canopy homolog 2 (MIR-interacting saposin-like protein) (Transmembrane protein 4) (Putative secreted protein ZSIG9)	Mus musculus (Mouse)	6.00E-17	SI.CL.2.cl.208.Contig1			putative ER protein	
SiJWG02ADE_pcr_1	E.SI.CL.64.cl.6455.Contig1		CNPY2_MOUSE	Protein canopy homolog 2 (MIR-interacting saposin-like protein) (Transmembrane protein 4) (Putative secreted protein ZSIG9)	Mus musculus (Mouse)	6.00E-17	SI.CL.2.cl.208.Contig1			putative ER protein	
SiJWH01BAA_pcr_2	E.SI.CL.64.cl.6455.Contig1		CNPY2_MOUSE	Protein canopy homolog 2 (MIR-interacting saposin-like protein) (Transmembrane protein 4) (Putative secreted protein ZSIG9)	Mus musculus (Mouse)	6.00E-17	SI.CL.2.cl.208.Contig1			putative ER protein	
SiJWB11AAJ_pcr_2	E.SI.CL.64.cl.6455.Contig1		CNPY2_MOUSE	Protein canopy homolog 2 (MIR-interacting saposin-like protein) (Transmembrane protein 4) (Putative secreted protein ZSIG9)	Mus musculus (Mouse)	6.00E-17					
SiJWF05ABR_pcr_2	E.SI.CL.65.cl.6501.Contig1		ENPL_PIG	Endoplasmin (Heat shock protein 90 kDa beta member 1) (94 kDa glucose regulated protein) (GRP94) (gp96 homolog) (98 kDa protein kinase) (PPK98) (ppk98)	Sus scrofa (Pig)	2.00E-74	SiJWF05ABR.scf				
SiJWA11BAJ_pcr_2	E.SI.CL.65.cl.6515.Contig1		ALAT2_XENLA	Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamic--pyruvic transaminase 2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2)	Xenopus laevis (African clawed frog)	1.00E-154	SI.CL.9.cl.996.Contig1	Translation	Ribosomal	probable Alanine aminotransferase	
SiJWE09ACU_pcr_2	E.SI.CL.65.cl.6515.Contig1		ALAT2_XENLA	Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamic--pyruvic transaminase 2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2)	Xenopus laevis (African clawed frog)	1.00E-154	SI.CL.9.cl.996.Contig1	Translation	Ribosomal	probable Alanine aminotransferase	
SiJWG02BBJ_pcr_1	E.SI.CL.65.cl.6515.Contig1		ALAT2_XENLA	Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamic--pyruvic transaminase 2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2)	Xenopus laevis (African clawed frog)	1.00E-154	SI.CL.9.cl.996.Contig1	Translation	Ribosomal	probable Alanine aminotransferase	
SiJWG12ADL_pcr_1	E.SI.CL.65.cl.6515.Contig1		ALAT2_XENLA	Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamic--pyruvic transaminase 2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2)	Xenopus laevis (African clawed frog)	1.00E-154	SI.CL.9.cl.996.Contig1	Translation	Ribosomal	probable Alanine aminotransferase	
SiJWD11ACU_pcr_2	E.SI.CL.65.cl.6515.Contig1		ALAT2_XENLA	Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamic--pyruvic transaminase 2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2)	Xenopus laevis (African clawed frog)	1.00E-154					
SiJWC08AAM_pcr_2	E.SI.CL.65.cl.6517.Contig1	*	PSA2_DROME	Proteasome subunit alpha type-2 (EC 3.4.25.1) (Proteasome 25 kDa subunit) (PROS-Dm25)	Drosophila melanogaster (Fruit fly)	1.00E-108	SI.CL.24.cl.2429.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type	
SiJWH11ADQ_pcr_1	E.SI.CL.65.cl.6517.Contig1	*	PSA2_DROME	Proteasome subunit alpha type-2 (EC 3.4.25.1) (Proteasome 25 kDa subunit) (PROS-Dm25)	Drosophila melanogaster (Fruit fly)	1.00E-108	SI.CL.24.cl.2429.Contig1	Protein processing	Proteosome	Proteasome subunit alpha type	
SiJWA09AAS_pcr_2	E.SI.CL.65.cl.6519.Contig1						NA	SiJWA09AAS.scf			
SiJWD04BBG_pcr_1	E.SI.CL.65.cl.6519.Contig1						NA	SI.CL.10.cl.1081.Contig1			
SiJWD09ABE_pcr_2	E.SI.CL.65.cl.6519.Contig1						NA	SI.CL.10.cl.1081.Contig1			
SiJWC11BBH_pcr_1	E.SI.CL.65.cl.6549.Contig1		NDUA5_BOVIN	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5 (NADH-ubiquinone oxidoreductase 13 kDa-B subunit) (Complex I-13kD-B) (CI-13kD-B) (Complex I subunit B13)	Bos taurus (Bovine)	7.00E-30	SI.CL.8.cl.829.Contig1	Mitochondrion	ATP synthesis	NADH dehydrogenase subcomplex subunit	
SiJWA01CAN_pcr_1	E.SI.CL.65.cl.6598.Contig1		SPCS1_DROME	Signal peptidase complex subunit 1 (EC 3.4.-.-) (Microsomal signal peptidase 12 kDa subunit) (SPase 12 kDa subunit)	Drosophila melanogaster (Fruit fly)	5.00E-24	SI.CL.11.cl.1177.Contig1	Protein processing	Protease	Signal peptidase 12kDa subunit	
SiJWC06BDO_pcr_1	E.SI.CL.65.cl.6598.Contig1		SPCS1_DROME	Signal peptidase complex subunit 1 (EC 3.4.-.-) (Microsomal signal peptidase 12 kDa subunit) (SPase 12 kDa subunit)	Drosophila melanogaster (Fruit fly)	5.00E-24	SI.CL.11.cl.1177.Contig1	Protein processing	Protease	Signal peptidase 12kDa subunit	
SiJWG08ADZ_pcr_1	E.SI.CL.65.cl.6598.Contig1		SPCS1_DROME	Signal peptidase complex subunit 1 (EC 3.4.-.-) (Microsomal signal peptidase 12 kDa subunit) (SPase 12 kDa subunit)	Drosophila melanogaster (Fruit fly)	5.00E-24	SI.CL.11.cl.1177.Contig1	Protein processing	Protease	Signal peptidase 12kDa subunit	
SiJWF03BDL_pcr_1	E.SI.CL.66.cl.6611.Contig1		HCC1_PONAB	Nuclear protein Hcc-1	Pongo abelii (Sumatran orangutan)	2.00E-16	SI.CL.22.cl.2229.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential	
SiJWG10BCH_pcr_1	E.SI.CL.66.cl.6621.Contig1		PSH_DROME	Serine protease persephone (EC 3.4.21.-)	Drosophila melanogaster (Fruit fly)	2.00E-36	SI.CL.3.cl.314.Contig1	Protein processing	Protease	Serine protease	
SiJWG03ABR_pcr_2	E.SI.CL.66.cl.6624.Contig1					2.00E-12	SI.CL.1.cl.134.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential	
SiJWH04AAD_pcr_2	E.SI.CL.66.cl.6624.Contig1					2.00E-12	SI.CL.1.cl.134.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential	
SiJWA06ADA_pcr_1	E.SI.CL.67.cl.6707.Contig1		RT26_DROME	Probable 28S ribosomal protein S26, mitochondrial (S26mt) (MRP-S26)	Drosophila melanogaster (Fruit fly)	7.00E-30	SI.CL.23.cl.2320.SiJWA06ADA.scf	Mitochondrion	Translation	Probable 28S ribosomal protein S26 mitochondrial precursor	
SiJWF02ACH_pcr_2	E.SI.CL.67.cl.6715.Contig1		CC124_DANRE	Coiled-coil domain-containing protein 124	Danio rerio (Zebrafish) (Brachydanio rerio)	3.00E-27	SI.CL.10.cl.1071.Contig1	Uncharacterized	Other	conserved protein, unknown function	
SiJWG04BBI_pcr_1	E.SI.CL.70.cl.7068.Contig1						NA	SI.CL.24.cl.2471.Contig1			
SiJWD04CAG_pcr_1	E.SI.CL.74.cl.7443.Contig1						NA	SI.CL.36.cl.3604.Contig1			
SiJWD04CAG_pcr_2	E.SI.CL.74.cl.7443.Contig1						NA	SI.CL.36.cl.3604.Contig1			
SiJWF02BDD_pcr_1	E.SI.CL.76.cl.7682.Contig1						NA	SI.CL.36.cl.3668.Contig1			
SiJWG05BCY_pcr_1	E.SI.CL.76.cl.7682.Contig1						NA	SI.CL.36.cl.3668.Contig1			
SiJWF09BDD_pcr_1	E.SI.CL.77.cl.7725.Contig1		B0WFT9_CULQU	Putative uncharacterized protein	Culex quinquefasciatus (Southern house mosquito) (Culex pungens)	1.00E-06	SI.CL.6.cl.628.Contig1	Uncharacterized	Other	uncharacterized protein	
SiJWA03BDM_pcr_1	E.SiJWA03BDM2.scf						NA	SI.CL.27.cl.2735.Contig1			
SiJWA09BAF_pcr_1	E.SiJWA09BAF.scf						NA	SiJWA09BAF.scf			
SiJWC11ACS_pcr_1	E.SiJWC11ACS.scf					1.00E-13	SiJWC11ACS.scf	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential	
SiJWE01ADK_pcr_1	E.SiJWE01ADK.scf						NA				
SiJWE07BDR_pcr_1	E.SiJWE07BDR2.scf					5.00E-23	SI.CL.10.cl.1071.SiJWE07BDR2.scf	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential	
SiJWE11BCZ_pcr_1	E.SiJWE11BCZ2.scf						NA	SiJWE11BCZ2.scf			
SiJWF01ACL_pcr_2	E.SiJWF01ACL.scf						NA	SiJWF01ACL.scf			
SiJWF06BCA_pcr_1	E.SiJWF06BCA2.scf		BAT1_MOUSE	B(0,+)-type amino acid transporter 1 (B(0,+)AT) (Glycoprotein-associated amino acid transporter b0,+AT1) (Solute carrier family 7 member 9)	Mus musculus (Mouse)	5.00E-07	SiJWF06BCA2.scf			Cationic amino acid transporter, potential	
SiJWH01ACJ_pcr_2	E.SiJWH01ACJ_for_Lino.ab1						NA	SI.CL.0.cl.003.Contig1			

Supporting Table 2

List of all fire ant genes significantly downregulated for at least one of the following comparisons: 6h vs oh, 24h vs 6h, 24h vs oh. Some genes are significant according to multiple microarray clones.

downatOnepoint		Swissprot entry with highest similarity to each fire ant gene								
Clone	"gene" = Contig from March 2009 "E" assembly	Entry name	Protein names	Organism	evalue	Contig from Oct 2008 Fourmidable assembly	manual_GO1	manual_GO2	manual_annotation
SiJWG07AAK_pcr_2	E.SI.CL.00.cl.0048.Contig1				NA	SiJWG07AAK.scf			
SiJWD02ABH_pcr_2	E.SI.CL.00.cl.0074.Contig1	SCPDH_PONAB	Probable saccharopine dehydrogenase (EC 1.5.1.9)	Pongo abelii (Sumatran orangutan)	2.00E-26	SI.CL.13.cl.1312.Contig1			Saccharopine dehydrogenase domain-containing protein
SiJWG06ADY_pcr_1	E.SI.CL.00.cl.0074.Contig1	SCPDH_PONAB	Probable saccharopine dehydrogenase (EC 1.5.1.9)	Pongo abelii (Sumatran orangutan)	2.00E-26	SI.CL.13.cl.1312.Contig1			Saccharopine dehydrogenase domain-containing protein
SiJWG06ADY_pcr_new	E.SI.CL.00.cl.0074.Contig1	SCPDH_PONAB	Probable saccharopine dehydrogenase (EC 1.5.1.9)	Pongo abelii (Sumatran orangutan)	2.00E-26	SI.CL.13.cl.1312.Contig1			Saccharopine dehydrogenase domain-containing protein
SiJWH08CAM_pcr_1	E.SI.CL.00.cl.0074.Contig1	SCPDH_PONAB	Probable saccharopine dehydrogenase (EC 1.5.1.9)	Pongo abelii (Sumatran orangutan)	2.00E-26	SI.CL.13.cl.1312.Contig2			
SiJWD11BAJ_pcr_2	E.SI.CL.01.cl.0111.Contig1	MLP2_DROME	Muscle LIM protein Mlp84B	Drosophila melanogaster (Fruit fly)	1.00E-178	SI.CL.19.cl.1965.Contig1			no hit
SiJWE03BAQ_pcr_1	E.SI.CL.01.cl.0111.Contig1	MLP2_DROME	Muscle LIM protein Mlp84B	Drosophila melanogaster (Fruit fly)	1.00E-178	SI.CL.19.cl.1965.Contig1			no hit
SiJWE05BBR_pcr_1	E.SI.CL.01.cl.0111.Contig1	MLP2_DROME	Muscle LIM protein Mlp84B	Drosophila melanogaster (Fruit fly)	1.00E-178	SI.CL.19.cl.1965.SiJWE05BBR.scf			no hit
SiJWF12ADC_pcr_1	E.SI.CL.01.cl.0111.Contig1	MLP2_DROME	Muscle LIM protein Mlp84B	Drosophila melanogaster (Fruit fly)	1.00E-178	SI.CL.19.cl.1965.Contig1			no hit
SiJWC02BCE_pcr_1	E.SI.CL.01.cl.0111.SiJWC02BCE2.scf				NA	SI.CL.19.cl.1965.SiJWC02BCE2.scf			no hit
SiJWD08BCG_pcr_1	E.SI.CL.01.cl.0111.SiJWD08BCG2.scf				NA	SI.CL.19.cl.1965.SiJWD08BCG2.scf			no hit
SiJWH09BBA_pcr_1	E.SI.CL.01.cl.0111.SiJWH09BBA.scf				NA	SiJWH09BBA.scf			
SiJWB11ACB_pcr_2	E.SI.CL.01.cl.0121.Contig1	L2EFL_DROME	Protein lethal(2)essential for life (Protein Efl21)	Drosophila melanogaster (Fruit fly)	1.00E-11	SI.CL.3.cl.307.Contig1			
SiJWE05ACT_pcr_2	E.SI.CL.01.cl.0121.Contig1	L2EFL_DROME	Protein lethal(2)essential for life (Protein Efl21)	Drosophila melanogaster (Fruit fly)	1.00E-11	SI.CL.3.cl.307.Contig1			
SiJWH01ADV_pcr_1	E.SI.CL.01.cl.0121.Contig1	L2EFL_DROME	Protein lethal(2)essential for life (Protein Efl21)	Drosophila melanogaster (Fruit fly)	1.00E-11	SI.CL.3.cl.307.Contig1			
SiJWE08BBE_pcr_1	E.SI.CL.01.cl.0147.Contig1	GSTT1_BOVIN	Glutathione S-transferase theta-1 (EC 2.5.1.18) (GST class-theta-1)	Bos taurus (Bovine)	3.00E-41				
SiJWE09BDE_pcr_1	E.SI.CL.01.cl.0147.Contig1	GSTT1_BOVIN	Glutathione S-transferase theta-1 (EC 2.5.1.18) (GST class-theta-1)	Bos taurus (Bovine)	3.00E-41	SI.CL.27.cl.2754.Contig1			(GSTt) Glutathione S-transferase theta
SiJWE12BCU_pcr_1	E.SI.CL.01.cl.0147.Contig1	GSTT1_BOVIN	Glutathione S-transferase theta-1 (EC 2.5.1.18) (GST class-theta-1)	Bos taurus (Bovine)	3.00E-41	SI.CL.27.cl.2754.Contig2			(GSTt) Glutathione S-transferase theta
SiJWH06BBI_pcr_1	E.SI.CL.01.cl.0147.Contig1	GSTT1_BOVIN	Glutathione S-transferase theta-1 (EC 2.5.1.18) (GST class-theta-1)	Bos taurus (Bovine)	3.00E-41	SI.CL.27.cl.2754.SiJWH06BBI.scf			(GSTt) Glutathione S-transferase theta
SiJWB04BDZ_pcr_1	E.SI.CL.01.cl.0179.Contig1				NA	SI.CL.7.cl.757.Contig1			
SiJWG03ACL_pcr_2	E.SI.CL.01.cl.0179.Contig1				NA	SI.CL.7.cl.757.Contig1			
SiJWA09AAV_pcr_2	E.SI.CL.01.cl.0185.Contig1	B7QM49_IXOSC	Putative uncharacterized protein	Ixodes scapularis (Black-legged tick) (Deer	9.00E-24	SiJWA09AAV.scf			
SiJWH07AAA_pcr_2	E.SI.CL.03.cl.0317.Contig2				NA	SI.CL.4.cl.420.Contig3			
SiJWH04ADX_pcr_1	E.SI.CL.04.cl.0424.Contig1	SUCA_DROME	Succinyl-CoA ligase [GDP-forming] subunit alpha, mitochondrial (EC 6.2.1.-	Drosophila melanogaster (Fruit fly)	1.00E-110	SI.CL.0.cl.076.Contig1	Metabolism	TCA	Succinyl-CoA synthetase alpha subunit, probable
SiJWB10ACK_pcr_2	E.SI.CL.05.cl.0528.Contig1	ALF_DROME	Fructose-bisphosphate aldolase (EC 4.1.2.13)	Drosophila melanogaster (Fruit fly)	1.00E-137	SiJWB10ACK.scf	Metabolism	Glycogenesis	Fructose-bisphosphate aldolase
SiJWB06BDD_pcr_1	E.SI.CL.06.cl.0645.Contig2	ESTF_MYZPE	Esterase FE4 (EC 3.1.1.1) (Carboxylic-ester hydrolase)	Myzus persicae (Peach-potato aphid)	2.00E-97	SI.CL.26.cl.2605.Contig2	Signaling	JH related	Juvenile hormone esterase, putative
SiJWH01ADW_pcr_1	E.SI.CL.06.cl.0645.Contig4	ESTF_MYZPE	Esterase FE4 (EC 3.1.1.1) (Carboxylic-ester hydrolase)	Myzus persicae (Peach-potato aphid)	2.00E-06	SI.CL.26.cl.2605.Contig1	Signaling	JH related	Juvenile hormone esterase, putative
SiJWF02ABH_pcr_2	E.SI.CL.06.cl.0667.Contig1				NA	SiJWF02ABH.scf			
SiJWB12CAG_pcr_1	E.SI.CL.06.cl.0678.Contig1				NA	SiJWB12CAG2.scf			
SiJWG09BCC_pcr_1	E.SI.CL.08.cl.0829.Contig1				NA	SiJWG09BCC.tag5_G09_13.scf			
SiJWC12ADJ_pcr_1	E.SI.CL.09.cl.0921.Contig1	DHB4_RAT	Peroxisomal multifunctional enzyme type 2 (MFE-2) (D-bifunctional protein	Rattus norvegicus (Rat)	9.00E-84	SI.CL.2.cl.200.Contig1			dehydrogenase, probable
SiJWA10CAU_pcr_1	E.SI.CL.09.cl.0923.Contig1				NA	SiJWA10CAU2.scf			
SiJWD01AAO_pcr_2	E.SI.CL.09.cl.0973.Contig2	HYEP1_CTEFE	Juvenile hormone epoxide hydrolase 1 (EC 3.3.2.9) (Juvenile hormone epo	Ctenocephalides felis (Cat flea)	1.00E-134	SI.CL.11.cl.1173.SiJWD01.	Signaling	JH related	Juvenile hormone epoxide hydrolase
SiJWE08AAS_pcr_2	E.SI.CL.10.cl.1086.Contig1	GPDA_DROKA	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic (GPDH-C) (GPD	Drosophila kanekoi (Fruit fly)	1.00E-150	SI.CL.5.cl.534.Contig1			
SiJWF06ACR_pcr_2	E.SI.CL.10.cl.1086.Contig1	GPDA_DROKA	Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic (GPDH-C) (GPD	Drosophila kanekoi (Fruit fly)	1.00E-150	SI.CL.5.cl.534.Contig1			
SiJWA10BCV_pcr_1	E.SI.CL.11.cl.1125.Contig1	Q0IFI6_AEDAE	Putative uncharacterized protein	Aedes aegypti (Yellowfever mosquito) (Cul	7.00E-12	SI.CL.18.cl.1849.Contig1			
SiJWD10ABL_pcr_2	E.SI.CL.11.cl.1125.Contig1	Q0IFI6_AEDAE	Putative uncharacterized protein	Aedes aegypti (Yellowfever mosquito) (Cul	7.00E-12	SI.CL.18.cl.1849.Contig1			
SiJWA05ADO_pcr_1	E.SI.CL.11.cl.1155.Contig1				NA	SI.CL.25.cl.2551.Contig1			
SiJWE09ABH_pcr_2	E.SI.CL.11.cl.1167.Contig1	CREG1_MOUSE	Protein CREG1 (Cellular repressor of E1A-stimulated genes 1)	Mus musculus (Mouse)	1.00E-36	SI.CL.0.cl.031.Contig1	Cell cycle	Cyclins&al negative	Cellular repressor of E1A-stimulated genes
SiJWE10ABL_pcr_2	E.SI.CL.11.cl.1167.Contig1	CREG1_MOUSE	Protein CREG1 (Cellular repressor of E1A-stimulated genes 1)	Mus musculus (Mouse)	1.00E-36	SI.CL.0.cl.031.Contig1	Cell cycle	Cyclins&al negative	Cellular repressor of E1A-stimulated genes
SiJWC05AAP_pcr_2	E.SI.CL.12.cl.1288.Contig1	Q59HJ1_9HYME	Carboxylesterase	Athalia rosae (coleseed sawfly)	3.00E-08	SI.CL.3.cl.313.Contig1	Metabolism	Other	Carboxylesterase, potential
SiJWE03BDS_pcr_1	E.SI.CL.12.cl.1288.Contig1	Q59HJ1_9HYME	Carboxylesterase	Athalia rosae (coleseed sawfly)	3.00E-08	SI.CL.3.cl.313.Contig1	Metabolism	Other	Carboxylesterase, potential
SiJWF04ACK_pcr_2	E.SI.CL.12.cl.1288.Contig1	Q59HJ1_9HYME	Carboxylesterase	Athalia rosae (coleseed sawfly)	0.00000003	SI.CL.3.cl.313.Contig1	Metabolism	Other	Carboxylesterase, potential
SiJWG03CAA_pcr_1	E.SI.CL.12.cl.1288.Contig1	Q59HJ1_9HYME	Carboxylesterase	Athalia rosae (coleseed sawfly)	3.00E-08	SI.CL.3.cl.313.Contig1	Metabolism	Other	Carboxylesterase, potential
SiJWH11BCV_pcr_1	E.SI.CL.13.cl.1325.Contig1	DUT_RAT	Deoxyuridine 5'-triphosphate nucleotidohydrolase (dUTPase) (EC 3.6.1.23)	Rattus norvegicus (Rat)	1.00E-46	SI.CL.1.cl.127.Contig1			Deoxyuridine 5'-triphosphate nucleotidohydrolase
SiJWD11ADA_pcr_1	E.SI.CL.16.cl.1626.Contig1				NA	SiJWD11ADA.scf			
SiJWC03AAL_pcr_2	E.SI.CL.16.cl.1659.Contig1	IF4A2_RAT	Eukaryotic initiation factor 4A-II (eIF-4A-II) (eIF4A-II) (EC 3.6.1.-) (ATP-de	Rattus norvegicus (Rat)	1.00E-170	SI.CL.10.cl.1001.Contig2	Translation	Elongation/Initiation/Other	Translation initiation factor 4A (Eif4a)
SiJWC03AAZ_pcr_1	E.SI.CL.16.cl.1659.Contig1	IF4A2_RAT	Eukaryotic initiation factor 4A-II (eIF-4A-II) (eIF4A-II) (EC 3.6.1.-) (ATP-de	Rattus norvegicus (Rat)	1.00E-170	SI.CL.10.cl.1001.Contig2	Translation	Elongation/Initiation/Other	Translation initiation factor 4A (Eif4a)
SiJWC03AAZ_pcr_2	E.SI.CL.16.cl.1659.Contig1	IF4A2_RAT	Eukaryotic initiation factor 4A-II (eIF-4A-II) (eIF4A-II) (EC 3.6.1.-) (ATP-de	Rattus norvegicus (Rat)	1.00E-170	SI.CL.10.cl.1001.Contig2	Translation	Elongation/Initiation/Other	Translation initiation factor 4A (Eif4a)
SiJWD08BAA_pcr_2	E.SI.CL.16.cl.1659.Contig1	IF4A2_RAT	Eukaryotic initiation factor 4A-II (eIF-4A-II) (eIF4A-II) (EC 3.6.1.-) (ATP-de	Rattus norvegicus (Rat)	1.00E-170	SI.CL.10.cl.1001.Contig2	Translation	Elongation/Initiation/Other	Translation initiation factor 4A (Eif4a)
SiJWA05BAV_pcr_1	E.SI.CL.16.cl.1659.Contig1	IF4A2_RAT	Eukaryotic initiation factor 4A-II (eIF-4A-II) (eIF4A-II) (EC 3.6.1.-) (ATP-de	Rattus norvegicus (Rat)	1.00E-170	SI.CL.10.cl.1001.Contig1			
SiJWB08BDP_pcr_1	E.SI.CL.16.cl.1659.Contig2	Q4W7I9_9NEOP	Hypotheical protein (Fragment)	Nasutitermes takasagoensis	2.00E-06	SI.CL.10.cl.1001.Contig2	Translation	Elongation/Initiation/Other	Translation initiation factor 4A (Eif4a)
SiJWE03BBD_pcr_1	E.SI.CL.16.cl.1659.Contig2	Q4W7I9_9NEOP	Hypotheical protein (Fragment)	Nasutitermes takasagoensis	2.00E-06	SI.CL.10.cl.1001.Contig2	Translation	Elongation/Initiation/Other	Translation initiation factor 4A (Eif4a)
SiJWF09ACU_pcr_2	E.SI.CL.16.cl.1659.Contig2	Q4W7I9_9NEOP	Hypotheical protein (Fragment)	Nasutitermes takasagoensis	2.00E-06	SI.CL.10.cl.1001.Contig2	Translation	Elongation/Initiation/Other	Translation initiation factor 4A (Eif4a)
SiJWE03ABR_pcr_2	E.SI.CL.17.cl.1705.Contig3	CP4C1_BLADI	Cytochrome P450 4C1 (EC 1.14.14.1) (CYPIVC1)	Blaberus discoidalis (Tropical cockroach)	1.00E-107	SI.CL.13.cl.1348.Contig1	Signaling	JH related	Cytochrome P450
SiJWF11AAV_pcr_2	E.SI.CL.17.cl.1766.Contig2				NA	SI.CL.12.cl.1234.Contig1			
SiJWG12ABM_pcr_2	E.SI.CL.18.cl.1854.Contig1	EID3_RAT	EP300-interacting inhibitor of differentiation 3 (EID-3) (EID-1-like inhibitor	Rattus norvegicus (Rat)	5.00E-12	SiJWG12ABM.scf			
SiJWH06CAD_pcr_1	E.SI.CL.18.cl.1884.Contig1	YRBE_BACSU	Uncharacterized oxidoreductase yrbE (EC 1.-.-.-)	Bacillus subtilis	1.00E-46	SI.CL.0.cl.088.Contig1			Inositol 2-dehydrogenase, potential
SiJWB07BAV_pcr_1	E.SI.CL.20.cl.2011.Contig1	UCP3_BOVIN	Mitochondrial uncoupling protein 3 (UCP 3)	Bos taurus (Bovine)	5.00E-19	SiJWB07BAV.scf			
SiJWB05ADW_pcr_1	E.SI.CL.20.cl.2055.Contig1				NA	SI.CL.17.cl.1799.Contig1			
SiJWA03ABO_pcr_2	E.SI.CL.20.cl.2080.Contig1				NA	SI.CL.29.cl.2990.Contig1			
SiJWB12CAC_pcr_1	E.SI.CL.20.cl.2080.Contig1				NA	SI.CL.29.cl.2990.Contig1			

SiJWC11BBR_pcr_1	E.SI.CL.21.cl.2147.Contig1	GD_DROME	Serine protease gd (EC 3.4.21.-) (Protein gastrulation defective)	Drosophila melanogaster (Fruit fly)	4.00E-14	SiJWC11BBR.scf	Protein processing	Protease	Trypsin, putative
SiJWA11BBV_pcr_1	E.SI.CL.21.cl.2168.Contig1	CP4C1_BLADI	Cytochrome P450 4C1 (EC 1.14.14.1) (CYPIVC1)	Blaberus discoidalis (Tropical cockroach)	1.00E-104				
SiJWE11ABE_pcr_2	E.SI.CL.21.cl.2168.Contig1	CP4C1_BLADI	Cytochrome P450 4C1 (EC 1.14.14.1) (CYPIVC1)	Blaberus discoidalis (Tropical cockroach)	1.00E-104				
SiJWH04CAI_pcr_1	E.SI.CL.21.cl.2168.Contig2	CP4C1_BLADI	Cytochrome P450 4C1 (EC 1.14.14.1) (CYPIVC1)	Blaberus discoidalis (Tropical cockroach)	1.00E-115	SI.CL.32.cl.3224.Contig1	Signaling	JH related	Cytochrome P450
SiJWA07ACB_pcr_2	E.SI.CL.22.cl.2235.Contig1	CSN4_MOUSE	COP9 signalosome complex subunit 4 (Signalosome subunit 4) (SGN4) (JAI	Mus musculus (Mouse)	1.00E-175	SI.CL.0.cl.030.Contig1	Mitochondrion	Other	mitochondrial porin
SiJWA11ACT_pcr_2	E.SI.CL.22.cl.2270.Contig1				NA	SiJWA11ACT.scf			
SiJWF04CAQ_pcr_1	E.SI.CL.22.cl.2298.Contig1				NA	SI.CL.31.cl.3165.Contig1			
SiJWG12ACG_pcr_2	E.SI.CL.23.cl.2310.Contig1				NA	SI.CL.1.cl.126.Contig1			
SiJWC03BAW_pcr_1	E.SI.CL.23.cl.2361.Contig1				NA	SI.CL.2.cl.295.Contig1			
SiJWD09CAD_pcr_1	E.SI.CL.24.cl.2428.Contig1	SAP_HUMAN	Proactivator polypeptide [Cleaved into: Saposin-A (Protein A); Saposin-B-V	Homo sapiens (Human)	4.00E-39	SI.CL.6.cl.627.Contig1	Lysosome		saposin
SiJWD10BAP_pcr_1	E.SI.CL.24.cl.2428.Contig1	SAP_HUMAN	Proactivator polypeptide [Cleaved into: Saposin-A (Protein A); Saposin-B-V	Homo sapiens (Human)	4.00E-39	SI.CL.6.cl.627.Contig1	Lysosome		saposin
SiJWF05ABD_pcr_2	E.SI.CL.24.cl.2428.Contig1	SAP_HUMAN	Proactivator polypeptide [Cleaved into: Saposin-A (Protein A); Saposin-B-V	Homo sapiens (Human)	4.00E-39	SI.CL.6.cl.627.Contig1	Lysosome		saposin
SiJWF08CAT_pcr_1	E.SI.CL.24.cl.2428.Contig1	SAP_HUMAN	Proactivator polypeptide [Cleaved into: Saposin-A (Protein A); Saposin-B-V	Homo sapiens (Human)	4.00E-39	SI.CL.6.cl.627.Contig1	Lysosome		saposin
SiJWG01CAM_pcr_1	E.SI.CL.25.cl.2569.Contig1	CAH2_TRIHK	Carbonic anhydrase 2 (EC 4.2.1.1) (Carbonic anhydrase II) (CA-II) (Carbor	Triblodon hakonensis (Japanese dace)	9.00E-28	SI.CL.35.cl.3550.Contig1			Carbonic anhydrase
SiJWD06BBY_pcr_1	E.SI.CL.25.cl.2599.Contig1	GST1_BLAGE	Glutathione S-transferase (EC 2.5.1.18) (GST class-sigma) (Major allergen	Blattella germanica (German cockroach)	1.00E-45	SI.CL.22.cl.2206.Contig1			(GSTS1)Glutathione S-transferase
SiJWD05BCB_pcr_1	E.SI.CL.26.cl.2629.Contig1	YG015_MOUSE	Leucine-rich repeat and death domain-containing protein LOC401387 homc	Mus musculus (Mouse)	2.00E-17	SiJWD05BCB2.scf	Uncharacterized	Other	leucine rich repeat protein
SiJWH03BDE_pcr_1	E.SI.CL.26.cl.2692.Contig1	A2AXC1_TRICA	Gustatory receptor candidate 59	Tribolium castaneum (Red flour beetle)	2.00E-06	SI.CL.2.cl.242.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWA05ACX_pcr_1	E.SI.CL.27.cl.2784.Contig1	TRIPB_HUMAN	Thyroid receptor-interacting protein 11 (TRIP-11) (Golgi-associated microtı	Homo sapiens (Human)	1.00E-14	SI.CL.0.cl.046.Contig1			
SiJWB01BBF_pcr_1	E.SI.CL.27.cl.2784.Contig2				NA	SI.CL.0.cl.046.Contig1			
SiJWH03BCH_pcr_1	E.SI.CL.28.cl.2805.Contig1	ARP6_CHICK	Actin-related protein 6 (gArp6) (gARPX)	Gallus gallus (Chicken)	1.00E-119	SiJWH03BCH2.scf	Cytoskeleton/Matrix		actin-related protein
SiJWB07ACS_pcr_1	E.SI.CL.28.cl.2817.Contig1	CP4C1_BLADI	Cytochrome P450 4C1 (EC 1.14.14.1) (CYPIVC1)	Blaberus discoidalis (Tropical cockroach)	2.00E-45	SiJWB07ACS.scf	Signaling	JH related	Cytochrome P450
SiJWB07ACS_pcr_2	E.SI.CL.28.cl.2817.Contig1	CP4C1_BLADI	Cytochrome P450 4C1 (EC 1.14.14.1) (CYPIVC1)	Blaberus discoidalis (Tropical cockroach)	2.00E-45	SiJWB07ACS.scf	Signaling	JH related	Cytochrome P450
SiJWE11CAR_pcr_1	E.SI.CL.28.cl.2817.Contig3	CP4C1_BLADI	Cytochrome P450 4C1 (EC 1.14.14.1) (CYPIVC1)	Blaberus discoidalis (Tropical cockroach)	7.00E-40	SI.CL.2.cl.249.Contig1	Signaling	JH related	Cytochrome P450
SiJWB07BDA_pcr_1	E.SI.CL.29.cl.2905.Contig1	CA085_DANRE	Uncharacterized protein C1orf85 homolog	Danio rerio (Zebrafish) (Brachydanio rerio)	2.00E-27	SI.CL.31.cl.3198.Contig1			
SiJWD03CAC_pcr_1	E.SI.CL.29.cl.2926.Contig1				NA	SI.CL.42.cl.4258.Contig1			
SiJWA01BAO_pcr_1	E.SI.CL.29.cl.2964.Contig1	GTPB2_HUMAN	GTP-binding protein 2	Homo sapiens (Human)	1.00E-68	SI.CL.12.cl.1205.Contig1	Translation	Elongation/Initiation/Other	GTP binding protein
SiJWE03AAQ_pcr_2	E.SI.CL.29.cl.2964.Contig1	GTPB2_HUMAN	GTP-binding protein 3	Homo sapiens (Human)	1.00E-68	SI.CL.12.cl.1205.Contig1	Translation	Elongation/Initiation/Other	GTP binding protein
SiJWG09BBV_pcr_1	E.SI.CL.29.cl.2964.Contig1	GTPB2_HUMAN	GTP-binding protein 4	Homo sapiens (Human)	1.00E-68	SI.CL.12.cl.1205.Contig1	Translation	Elongation/Initiation/Other	GTP binding protein
SiJWH01BCX_pcr_1	E.SI.CL.30.cl.3037.Contig1				NA	SI.CL.42.cl.4203.Contig1			
SiJWE06CAK_pcr_1	E.SI.CL.30.cl.3092.Contig1				NA	SiJWE06CAK2.scf			
SiJWB12ACV_pcr_2	E.SI.CL.31.cl.3171.Contig1	RRBP1_CANFA	Ribosome-binding protein 1 (180 kDa ribosome receptor) (RRp)	Canis familiaris (Dog)	6.00E-06	SI.CL.17.cl.1726.Contig1			
SiJWA08BDD_pcr_1	E.SI.CL.31.cl.3179.Contig1				NA	SI.CL.36.cl.3650.Contig1			
SiJWG05ACW_pcr_2	E.SI.CL.31.cl.3185.Contig1				NA	SI.CL.10.cl.1030.SiJWG05ACW.scf			
SiJWC12ABJ_pcr_2	E.SI.CL.31.cl.3187.Contig1				NA	SI.CL.14.cl.1410.Contig1			
SiJWF11ACN_pcr_2	E.SI.CL.32.cl.3204.Contig1				NA	SI.CL.8.cl.827.Contig1			
SiJWH05ABX_pcr_2	E.SI.CL.32.cl.3282.Contig1				NA	SI.CL.7.cl.775.Contig1			
SiJWD02ADC_pcr_1	E.SI.CL.35.cl.3531.Contig1				NA	SiJWD02ADC.scf			
SiJWG01BBT_pcr_1	E.SI.CL.35.cl.3555.Contig1				NA	SiJWG01BBT.scf			
SiJWB01BCV_pcr_1	E.SI.CL.36.cl.3601.Contig1	MOSC2_MOUSE	MOSC domain-containing protein 2, mitochondrial (EC 1.-.-.-)	Mus musculus (Mouse)	3.00E-30	SI.CL.10.cl.1033.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWE01AAA_pcr_1	E.SI.CL.36.cl.3601.Contig1	MOSC2_MOUSE	MOSC domain-containing protein 2, mitochondrial (EC 1.-.-.-)	Mus musculus (Mouse)	3.00E-30	SI.CL.10.cl.1033.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWE01AAA_pcr_2	E.SI.CL.36.cl.3601.Contig1	MOSC2_MOUSE	MOSC domain-containing protein 2, mitochondrial (EC 1.-.-.-)	Mus musculus (Mouse)	3.00E-30	SI.CL.10.cl.1033.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWG11ACW_pcr_2	E.SI.CL.36.cl.3601.Contig1	MOSC2_MOUSE	MOSC domain-containing protein 2, mitochondrial (EC 1.-.-.-)	Mus musculus (Mouse)	3.00E-30	SI.CL.10.cl.1033.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWB12ABQ_pcr_2	E.SI.CL.36.cl.3618.Contig1				NA	SiJWB12ABQ.scf			
SiJWB10BCM_pcr_1	E.SI.CL.36.cl.3638.Contig1				NA	SiJWB10BCM.scf			
SiJWG08BEA_pcr_1	E.SI.CL.37.cl.3796.Contig1				NA	SI.CL.30.cl.3069.Contig1			
SiJWE09BDC_pcr_1	E.SI.CL.38.cl.3829.Contig1				NA				
SiJWB02BDJ_pcr_1	E.SI.CL.38.cl.3858.Contig1	CPR1_DROME	Putative cysteine proteinase CG12163 (EC 3.4.22.-)	Drosophila melanogaster (Fruit fly)	5.00E-80	SI.CL.23.cl.2314.Contig1			
SiJWD06CAB_pcr_1	E.SI.CL.38.cl.3858.SiJWD06CAB.scf				NA	SI.CL.23.cl.2314.Contig1			
SiJWG11AEA_pcr_1	E.SI.CL.38.cl.3858.SiJWG11AEA.scf				NA	SI.CL.23.cl.2314.Contig1			
SiJWH02BBK_pcr_1	E.SI.CL.39.cl.3952.Contig1				NA	SI.CL.13.cl.1347.Contig1			
SiJWD09ADR_pcr_1	E.SI.CL.39.cl.3973.Contig1				NA	SiJWD09ADR.scf			no hit
SiJWA09ABO_pcr_2	E.SI.CL.40.cl.4052.Contig1	FA12_HUMAN	Coagulation factor XII (EC 3.4.21.38) (Hageman factor) (HAF) [Cleaved int	Homo sapiens (Human)	8.00E-22	SI.CL.1.cl.103.Contig1	Protein processing	Protease	Peptidase
SiJWH12CAK_pcr_1	E.SI.CL.40.cl.4085.Contig1				NA	SI.CL.11.cl.1112.Contig1			
SiJWG03AAP_pcr_2	E.SI.CL.42.cl.4225.Contig1	Q7PJV2_ANOGA	AGAP010147-PA	Anopheles gambiae (African malaria mosq	2.00E-09	SI.CL.14.cl.1445.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWG09BDZ_pcr_1	E.SI.CL.42.cl.4259.Contig1				NA	SI.CL.7.cl.761.Contig1			
SiJWF02BBL_pcr_1	E.SI.CL.42.cl.4272.Contig1	HR96_DROME	Nuclear hormone receptor HR96 (dHR96) (Nuclear receptor subfamily 1 gr	Drosophila melanogaster (Fruit fly)	2.00E-66	SI.CL.21.cl.2195.Contig1	Signaling	Other	Nuclear hormone receptor
SiJWD10BDW_pcr_2	E.SI.CL.42.cl.4273.Contig1				NA	SI.CL.32.cl.3239.Contig1			
SiJWH03AAJ_pcr_2	E.SI.CL.43.cl.4309.Contig1				NA	SiJWH03AAJ.scf			
SiJWF10ACU_pcr_2	E.SI.CL.43.cl.4363.Contig1				NA	SiJWF10ACU.scf			
SiJWH02ABT_pcr_2	E.SI.CL.43.cl.4380.Contig1				NA	SI.CL.15.cl.1583.SiJWH02ABT.scf			
SiJWA08BCH_pcr_1	E.SI.CL.43.cl.4393.Contig1	Q29DE0_DROPS	GA12758	Drosophila pseudoobscura pseudoobscura	1.00E-10	SI.CL.42.cl.4219.SiJWA08BCH2.scf			
SiJWC09BAY_pcr_1	E.SI.CL.44.cl.4419.Contig1	ERCC5_XENLA	DNA repair protein complementing XP-G cells homolog (EC 3.1.-.-) (Xerode	Xenopus laevis (African clawed frog)	7.00E-17	SiJWC09BAY2.scf	Cell cycle	DNA repair	Nucleotide excision repair protein
SiJWE05BDJ_pcr_1	E.SI.CL.45.cl.4546.Contig1	IDS_HUMAN	Iduronate 2-sulfatase (EC 3.1.6.13) (Alpha-L-iduronate sulfate sulfatase) (	Homo sapiens (Human)	3.00E-18	SI.CL.14.cl.1479.Contig1			Iduronate 2-sulfatase, potential
SiJWG05BCH_pcr_1	E.SI.CL.45.cl.4546.Contig1	IDS_HUMAN	Iduronate 2-sulfatase (EC 3.1.6.13) (Alpha-L-iduronate sulfate sulfatase) (	Homo sapiens (Human)	3.00E-18	SI.CL.14.cl.1479.Contig1			Iduronate 2-sulfatase, potential
SiJWB06BAH_pcr_2	E.SI.CL.47.cl.4720.Contig1	B4MJK6_DROWI	GK20853	Drosophila willistoni (Fruit fly)	2.00E-14	SI.CL.5.cl.530.Contig1	Uncharacterized	Hymenoptera	hymenoptera-specific protein, potential
SiJWA12BBY_pcr_1	E.SI.CL.47.cl.4751.Contig1				NA	SI.CL.29.cl.2943.Contig1			
SiJWG06BBH_pcr_1	E.SI.CL.47.cl.4751.Contig1				NA	SI.CL.29.cl.2943.Contig1			
SiJWC05BDZ_pcr_1	E.SI.CL.48.cl.4866.Contig1	EF1A2_DROME	Elongation factor 1-alpha 2 (EF-1-alpha-2)	Drosophila melanogaster (Fruit fly)	0.00E+00	SI.CL.7.cl.734.Contig1	Translation	Elongation/Initiation/Other	Elongation factor-1alpha
SiJWF12ADW_pcr_1	E.SI.CL.48.cl.4866.Contig1	EF1A2_DROME	Elongation factor 1-alpha 2 (EF-1-alpha-2)	Drosophila melanogaster (Fruit fly)	0.00E+00	SI.CL.7.cl.734.Contig1	Translation	Elongation/Initiation/Other	Elongation factor-1alpha
SiJWE08BBN_pcr_1	E.SI.CL.48.cl.4871.Contig1				NA	SI.CL.30.cl.3062.Contig1			
SiJWD04ACB_pcr_2	E.SI.CL.49.cl.4958.Contig1				NA	SI.CL.3.cl.386.SiJWD04ACB.scf			
SiJWD03AAO_pcr_2	E.SI.CL.50.cl.5014.Contig1	CATA_DROME	Catalase (EC 1.11.1.6)	Drosophila melanogaster (Fruit fly)	0.00E+00				
SiJWE08AAM_pcr_2	E.SI.CL.50.cl.5014.SiJWE08AAM.sc	Q8I9W0_APILI	Catalase (EC 1.11.1.6)	Apis mellifera ligustica (Common honeybee	1.00E-07	SI.CL.40.cl.4085.Contig1			Catalase

SiJWC02BDB_pcr_1	E.SI.CL.50.cl.5054.Contig1					NA	SI.CL.37.cl.3727.Contig1			
SiJWD01AAB_pcr_1	E.SI.CL.50.cl.5091.Contig1	Q7Q298_ANOGA	AGAP003960-PA	Anopheles gambiae (African malaria mosq		1.00E-08	SI.CL.0.cl.079.SiJWD01AA	Protein processing	Protease	serine-type endopeptidase, potential
SiJWA11BCJ_pcr_1	E.SI.CL.51.cl.5100.SiJWA11BCJ3.scf					NA	SI.CL.38.cl.3827.SiJWA11BCJ3.scf			
SiJWG04BAC_pcr_1	E.SI.CL.51.cl.5122.Contig2	B4MWJ3_DROWI	GK19032	Drosophila willistoni (Fruit fly)		4.00E-23				
SiJWF12AAO_pcr_2	E.SI.CL.51.cl.5141.Contig1	LSD1_DROME	Lipid storage droplets surface-binding protein 1	Drosophila melanogaster (Fruit fly)		8.00E-70	SI.CL.5.cl.562.Contig1	Metabolism	Fatty acids met.	Lipid storage droplets surface-binding protein 1
SiJWB05BCO_pcr_1	E.SI.CL.51.cl.5142.Contig1					NA	SI.CL.38.cl.3848.Contig1			
SiJWC05ABV_pcr_2	E.SI.CL.51.cl.5172.Contig1					NA	SI.CL.15.cl.1594.Contig1			
SiJWH08ADH_pcr_1	E.SI.CL.52.cl.5228.Contig1					NA	SiJWH08ADH.scf			
SiJWE04ABG_pcr_2	E.SI.CL.53.cl.5317.Contig1					NA	SI.CL.16.cl.1691.Contig1			
SiJWE04ACP_pcr_2	E.SI.CL.53.cl.5317.Contig1					NA	SI.CL.16.cl.1691.Contig1			
SiJWF03ABK_pcr_2	E.SI.CL.53.cl.5317.Contig1					NA	SI.CL.16.cl.1691.Contig1			
SiJWA11BCQ_pcr_1	.53.cl.5317.SiJWA11BCQ_tag5_A11_01.scf					NA	SI.CL.16.cl.1691.Contig1			
SiJWD07BDK_pcr_1	:.SI.CL.53.cl.5317.SiJWD07BDK2.scf					NA	SI.CL.16.cl.1691.Contig1			
SiJWF03ACS_pcr_1	E.SI.CL.53.cl.5358.Contig1					NA	SiJWF03ACS.scf			
SiJWB06ABM_pcr_2	E.SI.CL.53.cl.5366.Contig1	CHSTB_HUMAN	Carbohydrate sulfotransferase 11 (EC 2.8.2.5) (Chondroitin 4-O-sulfotransf	Homo sapiens (Human)		6.00E-60	SI.CL.4.cl.439.Contig1			Chondroitin 4-sulfotransferase
SiJWE08BBK_pcr_1	E.SI.CL.54.cl.5467.Contig1	HGNAT_HUMAN	Heparan-alpha-glucosaminide N-acetyltransferase (EC 2.3.1.78) (Transmer	Homo sapiens (Human)		2.00E-59	SiJWE08BBK.scf			Heparan-alpha-glucosaminide N-acetyltransferase, possible
SiJWD11ADY_pcr_new	E.SI.CL.54.cl.5473.Contig1					NA	SiJWD11ADY.scf			
SiJWC12BBJ_pcr_1	E.SI.CL.55.cl.5548.Contig1					NA	SI.CL.16.cl.1603.Contig1			
SiJWE04ABV_pcr_2	E.SI.CL.55.cl.5548.Contig1					NA	SI.CL.16.cl.1603.Contig1			
SiJWG06BDO_pcr_1	E.SI.CL.55.cl.5548.Contig1					NA	SI.CL.16.cl.1603.Contig1			
SiJWA12AAD_pcr_2	E.SI.CL.55.cl.5594.Contig1	B4PGV6_DROYA	GE20094	Drosophila yakuba (Fruit fly)		1.00E-115	SI.CL.0.cl.001.Contig1	Uncharacterized	Other	conserved uncharacterized protein
SiJWB07BDM_pcr_1	E.SI.CL.56.cl.5643.Contig1	PUR9_RAT	Bifunctional purine biosynthesis protein PURH [Includes: Phosphoribosylam	Rattus norvegicus (Rat)		1.00E-130	SI.CL.6.cl.699.Contig1			Bifunctional purine biosynthesis protein
SiJWC07BCA_pcr_1	E.SI.CL.56.cl.5643.Contig1	PUR9_RAT	Bifunctional purine biosynthesis protein PURH [Includes: Phosphoribosylam	Rattus norvegicus (Rat)		1.00E-130	SI.CL.6.cl.699.Contig1			Bifunctional purine biosynthesis protein
SiJWF06AA1_pcr_1	E.SI.CL.56.cl.5643.Contig1	PUR9_RAT	Bifunctional purine biosynthesis protein PURH [Includes: Phosphoribosylam	Rattus norvegicus (Rat)		1.00E-130				
SiJWH04BDW_pcr_2	E.SI.CL.56.cl.5643.Contig1	PUR9_RAT	Bifunctional purine biosynthesis protein PURH [Includes: Phosphoribosylam	Rattus norvegicus (Rat)		1.00E-130	SI.CL.6.cl.699.Contig1			Bifunctional purine biosynthesis protein
SiJWH09ABI_pcr_2	E.SI.CL.56.cl.5663.Contig1					NA	SiJWH09ABI.scf			
SiJWA07ACZ_pcr_1	E.SI.CL.57.cl.5721.Contig1					NA	SI.CL.26.cl.2610.Contig1			
SiJWH08BBQ_pcr_1	E.SI.CL.58.cl.5883.Contig1					NA	SI.CL.31.cl.3142.Contig1			
SiJWG11ADL_pcr_1	E.SI.CL.59.cl.5911.Contig1	PRDM5_MOUSE	PR domain zinc finger protein 5 (PR domain-containing protein 5)	Mus musculus (Mouse)		7.00E-08	SI.CL.23.cl.2359.Contig1	Uncharacterized	Zinc finger	zinc finger domain protein
SiJWH04BCD_pcr_1	E.SI.CL.59.cl.5959.Contig1	EID3_RAT	EP300-interacting inhibitor of differentiation 3 (EID-3) (EID-1-like inhibitor	Rattus norvegicus (Rat)		8.00E-18				
SiJWG10CAT_pcr_1	E.SI.CL.60.cl.6031.Contig1					NA	SiJWG10CAT2.scf			
SiJWC06BCL_pcr_1	E.SI.CL.60.cl.6097.Contig1	F164C_MOUSE	UPF0418 protein FAM164C	Mus musculus (Mouse)		1.00E-25	SI.CL.5.cl.556.Contig1			
SiJWA12CAS_pcr_1	E.SI.CL.61.cl.6191.Contig1					NA	SI.CL.34.cl.3491.Contig1			
SiJWH10ADS_pcr_1	E.SI.CL.62.cl.6265.SiJWH10ADS.sc	B7QK64_IXOSC	Diacylglycerol O-acyltransferase, putative (EC 2.3.1.20)	Ixodes scapularis (Black-legged tick) (Deer		5.00E-08	SiJWH10ADS.scf	Uncharacterized	Other	conserved uncharacterized protein
SiJWB07CAN_pcr_1	E.SI.CL.62.cl.6276.Contig1	AL1A1_CHICK	Retinal dehydrogenase 1 (RALDH 1) (RalDH1) (EC 1.2.1.36) (Aldehyde deh	Gallus gallus (Chicken)		5.00E-81	SI.CL.39.cl.3936.Contig1			Aldehyde dehydrogenase
SiJWH05BCZ_pcr_1	E.SI.CL.63.cl.6307.Contig1					NA	SiJWH05BCZ2.scf			
SiJWB09AAD_pcr_2	E.SI.CL.63.cl.6321.Contig1	GPX4_CITSI	Probable phospholipid hydroperoxide glutathione peroxidase (PHGPx) (EC 1	Citrus sinensis (Sweet orange)		2.00E-45	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWD06AAW_pcr_2	E.SI.CL.63.cl.6321.Contig1	GPX4_CITSI	Probable phospholipid hydroperoxide glutathione peroxidase (PHGPx) (EC 1	Citrus sinensis (Sweet orange)		2.00E-45				
SiJWE03BCB_pcr_1	E.SI.CL.63.cl.6321.Contig1	GPX4_CITSI	Probable phospholipid hydroperoxide glutathione peroxidase (PHGPx) (EC 1	Citrus sinensis (Sweet orange)		2.00E-45	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWF06ABZ_pcr_2	E.SI.CL.63.cl.6321.Contig1	GPX4_CITSI	Probable phospholipid hydroperoxide glutathione peroxidase (PHGPx) (EC 1	Citrus sinensis (Sweet orange)		2.00E-45	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWG01ACS_pcr_1	E.SI.CL.63.cl.6321.Contig2					NA	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWG01ACS_pcr_2	E.SI.CL.63.cl.6321.Contig2					NA	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWC05BAM_pcr_1	E.SI.CL.63.cl.6321.SiJWC05BAM.sc	GPX4_MESCR	Probable phospholipid hydroperoxide glutathione peroxidase (PHGPx) (EC 1	Mesembryanthemum crystallinum (Commc		8.00E-16	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWC09BBI_pcr_1	E.SI.CL.63.cl.6321.SiJWC09BBI.scf					NA	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWG04ACD_pcr_2	E.SI.CL.63.cl.6321.SiJWG04ACD.sc	GPX2_CAEEL	Probable glutathione peroxidase R05H10.5 (EC 1.11.1.9)	Caenorhabditis elegans		3.00E-12	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWG09BDA_pcr_1	:.SI.CL.63.cl.6321.SiJWG09BDA2.scf					NA	SI.CL.1.cl.159.Contig1			(Gpx) Glutathione peroxidase
SiJWA02BBB_pcr_1	E.SI.CL.63.cl.6347.Contig1					NA	SI.CL.5.cl.567.Contig1			no hit
SiJWC03ABC_pcr_2	E.SI.CL.63.cl.6347.Contig1					NA	SI.CL.5.cl.567.Contig1			no hit
SiJWD03BCP_pcr_1	E.SI.CL.63.cl.6347.Contig1					NA	SI.CL.5.cl.567.Contig1			no hit
SiJWA12ABS_pcr_2	E.SI.CL.64.cl.6473.Contig1					NA	SI.CL.13.cl.1361.Contig1			
SiJWF09CAV_pcr_1	E.SI.CL.65.cl.6509.Contig1					NA	SI.CL.1.cl.180.Contig1			
SiJWH04ACB_pcr_2	E.SI.CL.65.cl.6509.Contig1					NA	SI.CL.1.cl.180.Contig1			
SiJWH10AAN_pcr_2	E.SI.CL.65.cl.6509.Contig1					NA	SI.CL.1.cl.180.Contig1			
SiJWH05ACI_pcr_2	E.SI.CL.65.cl.6509.Contig2					NA	SI.CL.1.cl.180.Contig1			
SiJWH11BCJ_pcr_1	E.SI.CL.65.cl.6509.Contig2					NA	SI.CL.1.cl.180.Contig1			
SiJWF07BBY_pcr_1	E.SI.CL.66.cl.6614.Contig1	Q7Q9Y6_ANOGA	AGAP004528-PA	Anopheles gambiae (African malaria mosq		1.00E-36	SiJWF07BBY.scf	Metabolism	Fatty acids	acyl-CoA acyltransferase, probable
SiJWB10ADS_pcr_1	E.SI.CL.71.cl.7140.Contig1					NA	SI.CL.25.cl.2574.Contig1			
SiJWG05ABU_pcr_2	E.SI.CL.77.cl.7729.Contig1					NA				
SiJWF09CAN_pcr_1	E.SI.CL.77.cl.7756.Contig1					NA	SI.CL.39.cl.3946.Contig1			
SiJWG12CAH_pcr_1	E.SI.CL.78.cl.7880.Contig1					NA				
SiJWA01BDX_pcr_1	E.SiJWA01BDX.scf					NA				
SiJWA02BCV_pcr_1	E.SiJWA02BCV.scf					NA	SiJWA02BCV.scf			
SiJWA04CAO_pcr_1	E.SiJWA04CAO2.scf					NA	SiJWA04CAO2.scf			
SiJWA12BDL_pcr_1	E.SiJWA12BDL2.scf	MAAI2_DROME	Probable maleylacetoacetate isomerase 2 (MAAI 2) (EC 5.2.1.2)	Drosophila melanogaster (Fruit fly)		1.00E-06				
SiJWB04BDJ_pcr_1	E.SiJWB04BDJ.scf					NA	SiJWB04BDJ.scf			
SiJWD01ABB_pcr_2	E.SiJWD01ABB.scf	MOCS3_DROAN	Adenylyltransferase and sulfurtransferase MOCS3 (Molybdenum cofactor sy	Drosophila ananassae (Fruit fly)		1.00E-45	SiJWD01ABB.scf			Mo-molybdopterin cofactor sulfurase, probable
SiJWD03CAB_pcr_1	E.SiJWD03CAB.scf					NA	SiJWD03CAB.scf			
SiJWD11ACQ_pcr_2	E.SiJWD11ACQ.scf					NA	SiJWD11ACQ.scf			
SiJWE02AAT_pcr_2	E.SiJWE02AAT.scf					NA	SiJWE02AAT.scf			
SiJWH08BCZ_pcr_1	E.SiJWH08BCZ2.scf					NA	SiJWH08BCZ2.scf			
SiJWH12BCR_pcr_1	E.SiJWH12BCR.scf					NA	SiJWH12BCR.scf			

Supporting Table 3

List of all fire ant genes significantly upregulated for at least one of the following comparisons: 6h vs 0h, 24h vs 6h, 24h vs 0h and also significantly higher in brains of honey bee queens than reproductive workers

				Swissprot entries with highest similarity to each fire ant gene		
	Fire ant gene	Bee gene	Putatively proteosomal	Entry name	Protein names	Organism
1	E.SI.CL.01.cl.0116.Contig1	GB16189	*	PSMD4_DROME	26S proteasome non-ATPase regulatory subunit 4 (26S proteasome regulatory subunit S5A) (Multiubiquitin chain-binding protein) (54 kDa subunit of mu particle) (p54)	Drosophila melanogaster (Fruit fly)
2	E.SI.CL.07.cl.0752.Contig1	GB16889		ESTF_MYZPE	Esterase FE4 (EC 3.1.1.1) (Carboxylic-ester hydrolase)	Myzus persicae (Peach-potato aphid)
3	E.SI.CL.11.cl.1152.Contig1	GB10411	*	PSA71_DROME	Proteasome subunit alpha type-7-1 (EC 3.4.25.1) (Proteasome 28 kDa subunit 1) (PROS-Dm28.1)	Drosophila melanogaster (Fruit fly)
4	E.SI.CL.11.cl.1156.Contig1	GB16965		PRDX5_MOUSE	Peroxiredoxin-5, mitochondrial (EC 1.11.1.15) (Prx-V) (Peroxisomal antioxidant enzyme) (PLP) (Thioredoxin reductase) (Thioredoxin peroxidase PMP20) (Antioxidant enzyme B166) (AOEB166) (Liver tissue 2D-page spot 2D-0014IV)	Mus musculus (Mouse)
5	E.SI.CL.17.cl.1789.Contig1	GB17251		NDKA_DROYA	Nucleoside diphosphate kinase (NDP kinase) (NDK) (EC 2.7.4.6) (Abnormal wing disks protein) (Fragment)	Drosophila yakuba (Fruit fly)
6	E.SI.CL.17.cl.1789.Contig2	GB17251		NDK_NEUCR	Nucleoside diphosphate kinase (NDP kinase) (NDK) (EC 2.7.4.6)	Neurospora crassa
7	E.SI.CL.19.cl.1988.Contig1	GB14191	*	PSB1_DROME	Proteasome subunit beta type-1 (EC 3.4.25.1) (Proteasome 26 kDa subunit)	Drosophila melanogaster (Fruit fly)
8	E.SI.CL.47.cl.4726.Contig1	GB15613		S61G1_GRYOR	Protein transport protein Sec61 subunit gamma	Gryllotalpa orientalis (Oriental mole cricket)
9	E.SI.CL.50.cl.5069.Contig1	GB19166		RAB1A_LYMST	Ras-related protein Rab-1A	Lymnaea stagnalis (Great pond snail)
10	E.SI.CL.60.cl.6036.Contig1	GB17986		PRAF1_HUMAN	Prenylated Rab acceptor protein 1 (PRA1 family protein 1)	Homo sapiens (Human)
11	E.SI.CL.65.cl.6515.Contig1	GB17641		ALAT2_XENLA	Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamic--pyruvic transaminase 2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2)	Xenopus laevis (African clawed frog)
12	E.SI.CL.65.cl.6517.Contig1	GB10390	*	PSA2_DROME	Proteasome subunit alpha type-2 (EC 3.4.25.1) (Proteasome 25 kDa subunit) (PROS-Dm25)	Drosophila melanogaster (Fruit fly)

Supporting Table 4

List of all fire ant genes significantly upregulated for at least one of the following comparisons: 6h vs 0h, 24h vs 6h, 24h vs 0h and also significantly upregulated after mating in honey bee queens

				Swissprot entries with highest similarity to each fire ant gene		
	Fire ant gene	Bee gene	Putatively proteasomal	Entry name	Protein names	Organism
1	E.SI.CL.01.cl.0142.Contig1	GB10521		RM14_RAT	39S ribosomal protein L14, mitochondrial (L14mt) (MRP-L14) (MRP-L32)	Rattus norvegicus (Rat)
2	E.SI.CL.06.cl.0679.Contig1	GB16413		TMCO1_RAT	Transmembrane and coiled-coil domains protein 1 (Meg-2-like protein)	Rattus norvegicus (Rat)
3	E.SI.CL.14.cl.1496.Contig1	GB18495		ARL2_HUMAN	ADP-ribosylation factor-like protein 2	Homo sapiens (Human)
4	E.SI.CL.17.cl.1732.Contig1	GB10351	*	PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7-I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)
5	E.SI.CL.19.cl.1988.Contig1	GB14191	*	PSB1_DROME	Proteasome subunit beta type-1 (EC 3.4.25.1) (Proteasome 26 kDa subunit)	Drosophila melanogaster (Fruit fly)
6	E.SI.CL.30.cl.3008.Contig2	GB13868		RT10_DROME	28S ribosomal protein S10, mitochondrial (S10mt) (MRP-S10)	Drosophila melanogaster (Fruit fly)
7	E.SI.CL.32.cl.3279.Contig1	GB16270		EXOS1_MOUSE	3'-5' exoribonuclease CSL4 homolog (EC 3.1.13.-) (Exosome component 1)	Mus musculus (Mouse)
8	E.SI.CL.33.cl.3387.Contig1	GB15477		RPAB1_PONAB	DNA-directed RNA polymerases I, II, and III subunit RPABC1 (RNA polymerases I, II, and III subunit ABC1) (DNA-directed RNA polymerase II subunit E) (RPB5)	Pongo abelii (Sumatran orangutan)
9	E.SI.CL.47.cl.4726.Contig1	GB15613		S61G1_GRYOR	Protein transport protein Sec61 subunit gamma	Gryllotalpa orientalis (Oriental mole cricket)
10	E.SI.CL.47.cl.4786.Contig1	GB15416		LACB2_XENLA	Beta-lactamase-like protein 2 (EC 3.-.-.-)	Xenopus laevis (African clawed frog)
11	E.SI.CL.50.cl.5004.Contig1	GB11482	*	PSA4_HUMAN	Proteasome subunit alpha type-4 (EC 3.4.25.1) (Proteasome component C9) (Macropain subunit C9) (Multicatalytic endopeptidase complex subunit C9) (Proteasome subunit L)	Homo sapiens (Human)
12	E.SI.CL.62.cl.6284.Contig1	GB18934		RPAB2_DROME	DNA-directed RNA polymerases I, II, and III subunit RPABC2 (RNA polymerases I, II, and III subunit ABC2) (RPB6)	Drosophila melanogaster (Fruit fly)
13	E.SI.CL.65.cl.6517.Contig1	GB10390	*	PSA2_DROME	Proteasome subunit alpha type-2 (EC 3.4.25.1) (Proteasome 25 kDa subunit) (PROS-Dm25)	Drosophila melanogaster (Fruit fly)

Supporting Table 5

List of all fire ant genes significantly upregulated for at least one of the following comparisons: 6h vs oh, 24h vs 6h, 24h vs oh and also significantly upregulated in *Anopheles gambiae* females in response to mating according to Vectorbase gene expression data

				Swissprot entries with highest similarity to each fire ant gene		
	Fire ant gene	Anopheles gene	Putatively proteosomal	Entry name	Protein names	Organism
1	E.SI.CL.01.cl.0123.Contig1	AGAP010010		sp|Q13454|TUSC3_HUMAN	Tumor suppressor candidate 3 (Protein N33)	Homo sapiens (Human)
2	E.SI.CL.07.cl.0752.Contig1	AGAP005837		sp|P35502|ESTF_MYZPE	Esterase FE4 (EC 3.1.1.1) (Carboxylic-ester hydrolase)	Myzus persicae (Peach-potato aphid)
3	E.SI.CL.09.cl.0936.Contig1	AGAP008816	*	sp|Q9Z2U1|PSA5_MOUSE	Proteasome subunit alpha type-5 (EC 3.4.25.1) (Proteasome zeta chain) (Macropain zeta chain) (Multicatalytic endopeptidase complex zeta chain)	Mus musculus (Mouse)
4	E.SI.CL.09.cl.0945.Contig1	AGAP010679		sp|Q9W2D6|TIM10_DROME	Mitochondrial import inner membrane translocase subunit Tim10	Drosophila melanogaster (Fruit fly)
5	E.SI.CL.11.cl.1156.Contig1	AGAP001325		sp|P99029|PRDX5_MOUSE	Peroxiredoxin-5, mitochondrial (EC 1.11.1.15) (Prx-V) (Peroxisomal antioxidant enzyme) (PLP) (Thioredoxin reductase) (Thioredoxin peroxidase PMP20) (Antioxidant enzyme B166) (AOEB166) (Liver tissue 2D-page spot 2D-0014IV)	Mus musculus (Mouse)
6	E.SI.CL.13.cl.1352.Contig1	AGAP010137		sp|Q9LZG0|ADK2_ARATH	Adenosine kinase 2 (AK 2) (EC 2.7.1.20) (Adenosine 5'-phosphotransferase 2)	Arabidopsis thaliana (Mouse-ear cress)
7	E.SI.CL.14.cl.1426.Contig1	AGAP007088		sp|P24367|PPIB_CHICK	Peptidyl-prolyl cis-trans isomerase B (PPIase) (Rotamase) (EC 5.2.1.8) (Cyclophilin B) (S-cyclophilin) (SCYLP)	Gallus gallus (Chicken)
8	E.SI.CL.16.cl.1654.Contig1	AGAP005861		sp|P23438|SSRB_CANFA	Translocon-associated protein subunit beta (TRAP-beta) (Signal sequence receptor subunit beta) (SSR-beta) (Glycoprotein 25H) (gp25H)	Canis familiaris (Dog)
9	E.SI.CL.17.cl.1732.Contig1	AGAP008837	*	sp|P40307|PSB2_RAT	Proteasome subunit beta type-2 (EC 3.4.25.1) (Proteasome component C7-I) (Macropain subunit C7-I) (Multicatalytic endopeptidase complex subunit C7-I)	Rattus norvegicus (Rat)
10	E.SI.CL.19.cl.1988.Contig1	AGAP004991	*	sp|P40304|PSB1_DROME	Proteasome subunit beta type-1 (EC 3.4.25.1) (Proteasome 26 kDa subunit)	Drosophila melanogaster (Fruit fly)
11	E.SI.CL.28.cl.2866.Contig1	AGAP011383		sp|Q9V9Z9|EBP2_DROME	Probable rRNA-processing protein EBP2 homolog	Drosophila melanogaster (Fruit fly)
12	E.SI.CL.29.cl.2944.Contig1	AGAP003016		sp|Q9XZ63|ARMET_DROME	ARMET-like protein	Drosophila melanogaster (Fruit fly)
13	E.SI.CL.32.cl.3212.Contig1	AGAP004412		sp|Q9H2P9|DPH5_HUMAN	Diphthine synthase (EC 2.1.1.98) (Diphthamide biosynthesis methyltransferase)	Homo sapiens (Human)
14	E.SI.CL.32.cl.3262.Contig1	AGAP007644		sp|A4FV84|MRT4_BOVIN	mRNA turnover protein 4 homolog	Bos taurus (Bovine)
15	E.SI.CL.32.cl.3299.Contig1	AGAP010440		sp|Q9VLV5|RUXE_DROME	Probable small nuclear ribonucleoprotein E (snRNP-E) (Sm protein E) (Sm-E) (SmE)	Drosophila melanogaster (Fruit fly)
16	E.SI.CL.37.cl.3779.Contig1	AGAP001973	*	sp|P25788|PSA3_HUMAN	Proteasome subunit alpha type-3 (EC 3.4.25.1) (Proteasome component C8) (Macropain subunit C8) (Multicatalytic endopeptidase complex subunit C8)	Homo sapiens (Human)
17	E.SI.CL.38.cl.3821.Contig1	AGAP011729		tr|Q17J48|Q17J48_AEDAE	Inosine-uridine preferring nucleoside hydrolase (Fragment)	Aedes aegypti (Yellowfever mosquito) (Culex aegypti)
18	E.SI.CL.44.cl.4466.Contig1	AGAP003183		sp|P55735|SEC13_HUMAN	Protein SEC13 homolog (SEC13-related protein) (SEC13-like protein 1)	Homo sapiens (Human)
19	E.SI.CL.45.cl.4597.Contig1	AGAP001311		sp|Q5U4Z8|RM20_XENLA	39S ribosomal protein L20, mitochondrial (L20mt) (MRP-L20)	Xenopus laevis (African clawed frog)
20	E.SI.CL.47.cl.4768.Contig1	AGAP001919		sp|Q922R8|PDIA6_MOUSE	Protein disulfide-isomerase A6 (EC 5.3.4.1) (Thioredoxin domain-containing protein 7)	Mus musculus (Mouse)
21	E.SI.CL.48.cl.4862.Contig1	AGAP008099		sp|Q6BBI8|UFC1_RAT	Ufm1-conjugating enzyme 1 (Ubiquitin-fold modifier-conjugating enzyme 1)	Rattus norvegicus (Rat)
22	E.SI.CL.49.cl.4982.Contig1	AGAP012407		sp|P54399|PDI_DROME	Protein disulfide-isomerase (PDI) (dPDI) (EC 5.3.4.1)	Drosophila melanogaster (Fruit fly)
23	E.SI.CL.50.cl.5004.Contig1	AGAP004960	*	sp|P25789|PSA4_HUMAN	Proteasome subunit alpha type-4 (EC 3.4.25.1) (Proteasome component C9) (Macropain subunit C9) (Multicatalytic endopeptidase complex subunit C9) (Proteasome subunit L)	Homo sapiens (Human)
24	E.SI.CL.50.cl.5057.Contig1	AGAP005228		sp|Q290F0|MED9_DROPS	Mediator of RNA polymerase II transcription subunit 9 (Mediator complex subunit 9)	Drosophila pseudoobscura pseudoobscura (Fruit fly)
25	E.SI.CL.52.cl.5214.Contig1	AGAP003981		sp|O08700|VPS45_RAT	Vacuolar protein sorting-associated protein 45 (rvps45)	Rattus norvegicus (Rat)
26	E.SI.CL.53.cl.5362.Contig1	AGAP010608		sp|P08081|CLCA_RAT	Clathrin light chain A (Lca)	Rattus norvegicus (Rat)
27	E.SI.CL.53.cl.5384.Contig1	AGAP002931		sp|P40945|ARF2_DROME	ADP-ribosylation factor 2 (dARF II)	Drosophila melanogaster (Fruit fly)
28	E.SI.CL.58.cl.5866.Contig1	AGAP011842		sp|Q5M8Y1|SPCS2_XENTR	Probable signal peptidase complex subunit 2 (EC 3.4.-.-) (Microsomal signal peptidase 25 kDa subunit) (SPase 25 kDa subunit)	Xenopus tropicalis (Western clawed frog) (Silurana tropicalis)
29	E.SI.CL.60.cl.6036.Contig1	AGAP003662		sp|Q9UI14|PRAF1_HUMAN	Prenylated Rab acceptor protein 1 (PRA1 family protein 1)	Homo sapiens (Human)
30	E.SI.CL.60.cl.6096.Contig1	AGAP001827		sp|Q60432|HYOU1_CRIGR	Hypoxia up-regulated protein 1 (150 kDa oxygen-regulated protein) (Orp150) (170 kDa glucose-regulated protein)	Cricetulus griseus (Chinese hamster)
31	E.SI.CL.62.cl.6225.Contig1	AGAP010718	*	sp|P28075|PSB5_RAT	Proteasome subunit beta type-5 (EC 3.4.25.1) (Proteasome epsilon chain) (Macropain epsilon chain) (Multicatalytic endopeptidase complex epsilon chain) (Proteasome subunit X) (Proteasome chain 6)	Rattus norvegicus (Rat)
32	E.SI.CL.63.cl.6380.Contig1	AGAP007670		sp|Q9V4Q8|RU2A_DROME	Probable U2 small nuclear ribonucleoprotein A' (U2 snRNP-A')	Drosophila melanogaster (Fruit fly)
33	E.SI.CL.64.cl.6455.Contig1	AGAP008615		sp|Q9QXT0|CNPY2_MOUSE	Protein canopy homolog 2 (MIR-interacting saposin-like protein) (Transmembrane protein 4) (Putative secreted protein ZSIG9)	Mus musculus (Mouse)
34	E.SI.CL.65.cl.6515.Contig1	AGAP000901		sp|Q6GM82|ALAT2_XENLA	Alanine aminotransferase 2 (ALT2) (EC 2.6.1.2) (Glutamic--pyruvic transaminase 2) (Glutamate pyruvate transaminase 2) (GPT 2) (Glutamic--alanine transaminase 2)	Xenopus laevis (African clawed frog)
35	E.SI.CL.66.cl.6611.Contig1	AGAP003229		sp|Q5R4V4|HCC1_PONAB	Nuclear protein Hcc-1	Pongo abelii (Sumatran orangutan)
36	E.SI.CL.67.cl.6707.Contig1	AGAP004481		sp|Q9VVN2|RT26_DROME	Probable 28S ribosomal protein S26, mitochondrial (S26mt) (MRP-S26)	Drosophila melanogaster (Fruit fly)

Microarray Data

Will be uploaded to the Gene Expression Omnibus database

6

THE OUTCOME OF A COMPETITION FOR REPRODUCTION BETWEEN FIRE ANT QUEENS IS LINKED TO THEIR GENE EXPRESSION PROFILES

by Yannick Wurm, John Wang, Kenneth G Ross, Laurent Keller

Kenneth G Ross & I collected the ants. I performed all experiments except final queen collections from competitions which were performed by Kenneth G Ross. I performed all analyses and wrote the manuscript. John Wang, Kenneth G Ross and Laurent Keller helped design the experiment and revise the manuscript.

ABSTRACT

In many animals living in groups the reproductive status of individuals is determined by their social status. In species with social hierarchies, the death of dominant individuals typically upheaves the social hierarchy and provides an opportunity for subordinate individuals to improve their social status. Such a phenomenon occurs in the monogyne form of the fire ant, *Solenopsis invicta*, where colonies typically contain a single wingless reproductive queen, thousands of workers and hundreds of winged non-reproductive virgin queens. Upon the death of the mother queen, many virgin queens shed their wings and initiate reproductive development instead of departing on a mating flight. Workers progressively execute almost all of them over the following weeks. The workers base their collective decision on pheromonal cues associated with the onset of reproductive development of the virgin queens which occurs after orphaning.

To examine the factors that determine which virgin queens are executed and which survive, we set up artificial competitions between queens from different colonies. Using microarrays, we found that queens from winning colonies showed higher mitochondrial as well as organ development activities 24 hours after orphaning than did queens from colonies that lost the competitions. Furthermore, queens from colonies where queens shed their wings faster after orphaning were more likely to survive competitions. Finally, higher wing shedding speed is linked to higher mitochondrial activity.

This article is *in preparation* for submission to *Molecular Ecology*.

6.1 BACKGROUND

Reproduction is monopolized by only a small number of individuals in many group-living animals. Which individuals reproduce can be determined by fights for dominance or territory, by seniority within the group, by genotype and by other factors (Solomon & French 1997; Keller 1993; Keller & Ross 1998; Keller & Reeve 1994). Although the social stimuli responsible for changes in reproductive hierarchies are well-documented in many animals (Solomon & French 1997), only a few studies have examined the molecular and physiological mechanisms linking such stimuli to changes in reproductive status. In the cichlid fish *Astatotilapia burtoni*, disappearance of the dominant male leads to rapid reactions in subordinate males, including dramatic changes in body coloration and behavior, growth of certain brain regions and increases in brain levels of *gonadotropin releasing hormone 1* and *early growth response factor 1* (Burmeister *et al.* 2005; White *et al.* 2002). Similarly, the transition from subordinate to breeder status in white-browed sparrow weavers is accompanied by changes in type of song, morphology of song-related brain areas, and an increase in levels of two hormone receptors and two synaptic proteins in a song-related brain area (Voigt *et al.* 2007). Changes in brain morphology also accompany the transition from subordinate to breeder status in naked mole rats (Holmes *et al.* 2007). While the previous studies provide valuable insight into the responses to social opportunities, they mostly focused on brain morphology and only few candidate genes.

Social insects provide excellent models for studying the mechanisms involved in reproductive competition (Roseler 1991; Roseler *et al.* 1984; Dietemann *et al.* 2006; Neumann *et al.* 2000; Keller 1993). In social bees, wasps and ants there is a clear division of labor with one or a few individuals monopolizing reproduction. Differences in reproductive roles are generally associated with tremendous physiological and behavioral modifications (Wilson 1971; Bourke & Franks 1995). This has led to many behavioral and hormone-based experiments including some in *Solenopsis invicta* fire ants that have even succeeded in isolating glands and compounds involved in maintaining social dominance hierarchies (Vargo & Laurel 1994; Vargo & Hulsey 2000; Vargo 1999; Brent & Vargo 2003; Vander Meer *et al.* 1980). Furthermore, it was determined that when orphaned fire ant workers are provided with two mated queens, they generally retain the more fertile queen and execute the other (Fletcher & Blum 1983), indicating that they choose the highest quality queen.

Investigating social life at a molecular-genetic scale has recently become possible with the development of new tools for social insects (The Honey Bee Genome Sequencing Consortium 2006; Wang *et al.* 2007; Wurm *et al.* 2009). Some of the first studies focused on identifying the genes involved in differences between reproductive and non-reproductive castes (Pereboom *et al.* 2005; Weil *et al.* 2009; Gräff *et al.*

2007), and others have investigated the link between social context and gene activity (Toth *et al.* 2007; Wang *et al.* 2008). However, still very little is known about the changes in gene expression associated with changes in reproductive roles (but see Chapter 5), nor about the molecular genetic-factors that determine which individuals are most able to improve their social standing .

The red imported fire ant, *Solenopsis invicta*, represents a particularly attractive model for studying the onset of competition between subordinate individuals. During the reproductive season, colonies of the monogyne form (single queen per colony) can produce hundreds or even thousands of young virgin daughter queens. These virgin queens spend the next few weeks building up fat reserves within the colony. Once they reach sexual maturity, they do not immediately become reproductive because such queens are perceived as threats to the dominant queen and executed by the workers (Vargo & Porter 1993; Vargo & Laurel 1994). Thus, virgin queens remain in the parental nest without reproducing until they participate in a mating flight and attempt to found their own colony. However, a remarkable alternative exists in *S. invicta* when the mother queen dies. During the days after orphaning, many young queens shed their wings and initiate reproductive development. This reproductive shift is accompanied by the emission of pheromonal signals to which nestmate queens and workers react. When virgin nestmate queens perceive such signals, they refrain from shedding their own wings and initiating reproductive development (Fletcher *et al.* 1983; Vargo 1999). When orphaned workers perceive pheromonal signals emitted by queens initiating reproductive development, they begin to tend to these queens (Fletcher & Blum 1981). However, if several queens produce signals associated with initiation of reproductive development, the workers will progressively execute almost all of them over the next few weeks (Fletcher & Blum 1983). The surviving virgin queen or queens are thus "elected" by workers to replace the mother queen. These queens are unmated and thus unable to replenish the colony's worker force. However, until the colony's workers have died out, the queens can lay thousands of haploid eggs that develop into haploid reproductive males (Tschinkel 2006).

The aim of this study was to examine some of the factors that determine the outcome of the competition for reproductive dominance that follows the loss of the dominant mother queen. We set up artificial competitions for reproductive dominance between queens from pairs of colonies to identify lineages of queens that are more and less successful. Using gene expression microarrays, we found that queens that are more successful have higher activities genes putatively involved in mitochondrial, organ development and cell differentiation activities. More successful queens were more likely to come from colonies where queens shed their wings faster than slower after orphaning, and we found a stronger link of mitochondrial activity to wing shedding than to success in competitions.

6.2 MATERIALS AND METHODS

Ant collection and rearing

Thirty monogyne (single reproductive queen) *S. invicta* fire ant colonies were collected in Athens and Lexington, GA, USA in June 2006. Seventeen of the colonies each containing at least 50 winged virgin queens and were subsequently used as queen donor colonies. The remaining 13 colonies were used as worker donor colonies. All colonies were returned to the laboratory and reared for one month under standard conditions (Jouvenaz *et al.* 1977). Queen and male destined brood were removed weekly. By the time our experiments were performed, the virgin queens were thus at least one month old. They were unlikely to be more than a few months old since several large-scale mating flights had already taken place that year before collection.

We determined that each study colony was of the monogyne social form using several lines of evidence. Nest shape, nest density and worker size distribution were used to make initial identifications of social form in the field (Shoemaker *et al.* 2006). Subsequently, monogyny was confirmed for each colony by the presence of a single, highly physogastric, wingless queen. Finally, the social form was further verified by electrophoretically detecting only the *B* but not the *b* allele of *Gp-9* in pooled samples of 20 workers from each colony (lack of *b* allele is diagnostic for monogyny in *S. invicta* in the USA (Ross 1997; Keller & Ross 1998; Krieger & Ross 2002; Shoemaker *et al.* 2006)).

Comparison of speed of wing shedding between colonies

Our aim was to set up competitions between queens that are faster and queens that are slower at shedding their wings after orphaning. There is no established was of determining wing shedding speed, thus we combined several measures of wing shedding speed into a single global measure. For this, we first simulated orphaning for 10 virgin queens from each of the 17 queen donor colonies and took note of how much time passed until each queen shed their wings. We used the resulting data to calculate several measures of wing shedding speed. The first measure was simply the amount of time that passed until two out of 10 orphaned queens had shed their wings. Another measure was the number of queens that had shed their wings within five days of orphaning. We calculated three additional measures of wing shedding speed (see Supporting Table 1). Each measure was used to rank the 17 colonies from fastest to slowest. Subsequently, the rankings of the colonies were combined by summation to obtain a global measure of wing shedding speed for each colony. This global measure was used to rank colonies from

slowest to fastest (see Supporting Table 2). One colony was discarded since it contained substantially fewer virgin queens than the other colonies, thus 16 ranked queen donor colonies remained.

The time required by individual queens for wing shedding was measured using orphaning simulations as follows: From the foraging area of each of the 17 queen donor colonies, we haphazardly collected 10 virgin queens that we individually placed with 2g of mixed workers and brood in 10cm diameter petri dishes. For five days every 6h, and thereafter once every 24h, each queen was inspected until she had shed three out of her four wings.

Competition setup

We used the previously established global ranking of the speed at which queens from the queen donor colonies to establish competitions between queens from pairs of faster and slower colonies.

First, we set up eight queenless receiving colonies. To make these receiving colonies as similar as possible, we mixed workers from 13 worker donor colonies. To do this, two days after all queens, males and queen- and male-destined brood had been removed, we combined workers and worker-destined brood from 13 colonies into one large tray. We then exhaled onto the combined workers for 20 minutes to provide possibly confusing olfactory stimuli to the ants. Subsequently, we placed the tray overnight at 4°C which slows down all worker activity and thus also aggressive behavior. The next day, the large tray was brought to room temperature and its contents equally distributed into eight rearing trays with nests, thus constituting eight receiving queenless colonies. We observed no worker mortality.

Next, from the seventeen queen donor colonies, we created eight pairs of donor colonies based on the global ranking of wing shedding speed. The fastest colony was paired with the colony that ranked 10th; the 2nd fastest was paired with the 11th; and so on. One colony was discarded. Within each pair of colonies, the faster colony ranked faster than the slower colony by the global ranking as well as each of the five individual rankings of wing shedding speed (see Supporting Tables 1 and 2).

Finally, we established a competition from each pair of colonies. For this we haphazardly collected 10 queens from the foraging area of a faster colony as well as from the paired slower colony. We immediately transferred these 20 queens into one of the queenless receiving colonies. This was repeated for each of the eight pairs of queen donor colonies. After 21 days, all surviving queens were collected. Their heads and thoraces were transferred to 1.5ml tubes for genotyping (below); their abdomens were dissected to compare reproductive status between surviving queens within each competition. We qualitatively compared ovarian development between surviving queens within each competition and designated the queen or queens with the strongest ovarian development as the winning queen or

queens. In several competitions, two or more queens had similarly strongly developed ovaries. In one competition, no surviving queens could be found.

Genotyping queens for identification of winning colonies

We used microsatellite genotyping to identify which queen donor colonies the surviving queens originated from. DNA was isolated from heads and thoraces of surviving queens using the Nucleospin Tissue kit (Macherey-Nagel). Additionally, DNA was isolated from 5 individuals that had been collected for microarrays (see below) from each queen donor colony. PCR amplification using fluorescently labeled primers for seven microsatellites (*Sol-11*, *Sol-20*, *Sol-42*, *Sol-49*, *Sol-55* (Krieger & Keller 1997) and *Sdag-C367* (Ascunce *et al.* 2009)) followed by sequencing on an ABI PRISM 3100 Genetic Analyzer and analysis using GeneMapper (Applied Biosystems) allowed definitive assignment of queens collected at the end of the competitions to one of the two queen donor colonies in each pair. In all but one competition in which the two winning queens did not originate from the same queen donor colony and one competition where no queens survived, we designated the winning colony as the queen donor colony from which the winning queen or queens originated.

RNA isolation and microarray hybridization

Our aim was to identify molecular markers that are likely to be responsible for post-orphaning differences in the likelihood of becoming a replacement queen (winners vs. losers) and that affect speed of wing shedding in virgin queens (faster vs. slower). However, by the time a queen can be assigned to these phenotype categories, the molecular changes responsible for that phenotype may have passed, and in case of losers, the queen is dead. Thus, we followed an indirect approach which consisted in comparing sisters of winning and losing queens on one hand, and sisters of faster and slower queens on the other hand. For this, examined gene expression using microarrays in orphaned virgin queens 24 hours after orphaning as follows.

From the foraging area of each of the sixteen queen donor colonies that were used for competitions, we haphazardly collected five virgin queens for which we simulated orphaning by transferring them into individual colony fractions made up of 2g of mixed workers and brood in 10cm diameter petri dishes. Twenty-four hours after simulated orphaning, all queens were individually flash-frozen with liquid nitrogen in tubes containing 1g of 1.4mm Zirconium Silicate beads (QuackenBush). Samples were

then stabilized until RNA isolation by the addition of 900μl of Trizol reagent (Invitrogen), followed by homogenization with a FastPrep instrument (MP Biomedicals) and storage at -80°C

Using the Trizol protocol, total RNA was isolated from all individuals. RNA was pooled from 5 individuals per queen donor colony and treated with DNA-free (Ambion). Subsequently, impurities were filtered away with MicroCon-30 spin columns (Millipore), and RNA quality was assessed on a 1% agarose gel prior to linear amplification using the MessageAmp II kit (Ambion). Amplified mRNA samples from the 16 donor colonies were labeled with Cy3 and hybridized against Cy5-labeled "common reference" RNA to microarrays made from 22,560 independent fire ant cDNA spots, and scanned as previously described (Wang *et al.* 2008). For all procedures, precautions including randomization of sample order were taken to avoid introducing unwanted biases.

Data Analysis

All statistical analyses were performed in R 2.9.0 (The R Development Core Team 2007). The effects of queen donor colony, competition, and winner/loser status on queen mortality using likelihood ratio tests on binomial generalized linear models using the lme4 package (Bates & Maechler online).

Microarray preprocessing

Median signal and background levels for each probe were extracted from scanned microarray images using Axon Genepix software. The 18,444 *Solenopsis invicta* cDNA spots yielding a single PCR band (Wang *et al.* 2007) and passing visual and automated inspection were used for normalization and quality control. The limma 2.16 package (Smyth 2004) in Bioconductor (The R Development Core Team 2007; Gentleman *et al.* 2004) was used for normexp background correction, print-tip loess normalization within arrays, and aquantile normalization between arrays. The arrayQualityMetrics package (Kauffmann *et al.* 2009) and custom R scripts were used for quality control.

Sequence information and annotation

Microarray clone sequences (Wang *et al.* 2007) were combined with two runs of 454 cDNA sequencing (Y. Wurm, D. Hahn and DD. Shoemaker, DH and DDS are at USDA-ARS, Gainesville, unpublished). High quality sequence information was thus available for 16,227 *S. invicta* cDNA clones. The cDNA clones were bioinformatically assigned to clusters based on sequence information and reciprocal blastn alignment scores (see (Wurm *et al.* 2009)). For each of 6,116 clusters, each putatively representing a

single genes or gene family, we retained for analysis the cDNA spot that showed the lowest amount of noise in a previous study with the same batch of microarrays (Chapter 5).

The sequences representing each cluster were aligned against the non-redundant protein database (EMBL release 99) using NCBI BLASTX 2.2.16 (E-value $<10^{-5}$). Subsequently, Gene Ontology (GO) (Ashburner *et al.* 2000) annotations were inferred as previously described (Wurm *et al.* 2009).

Detection of differently expressed gene sets

To identify gene expression differences between queens from winning and losing colonies, the limma package was used to construct a model incorporating fixed effects for winner-loser differences and labeling dye, while competitions were used as blocks (random effect). After bayesian fitting of the model, microarray clones were ranked by probability of differential expression. These rankings were used for gene set enrichment analysis (Subramanian *et al.* 2005) using the Kolmogorov-Smirnov-like test as implemented with the "elim" method in the topGO package (Alexa *et al.* 2006). In brief, this method evaluates whether genes with of specific gene sets tend to occur toward the top of the ranked list of genes.

To identify gene expression differences between queens from colonies where queens shed their wings faster and slower, the limma package was used to construct a model incorporating fixed effects for faster-slower differences and labeling dye, while competitions were used as blocks (random effect). Gene set enrichment analysis was conducted as above.

6.3 RESULTS

Surviving queens and identification of winning colonies

Twenty-eight of the 160 queens that participated in competitions were still alive after three weeks. Twenty-three of these surviving queens had shed their wings.

The genotypes of the surviving queens clearly identified them as coming from one of the two queen donor colonies that participated in their respective competition (Figure 6.1). In three competitions, all surviving queens were from a single colony which we thus designated as the "winning" colony. In one competition, no queens survived (competition *VI'* in Figure 6.1). In the remaining four competitions, queens from both competing colonies survived. Based on qualitative comparison of the ovarian development of surviving queens within each competition, we could identify a "winning" and "losing" queen donor colony in three of these four competitions. Surviving queens from the remaining competition

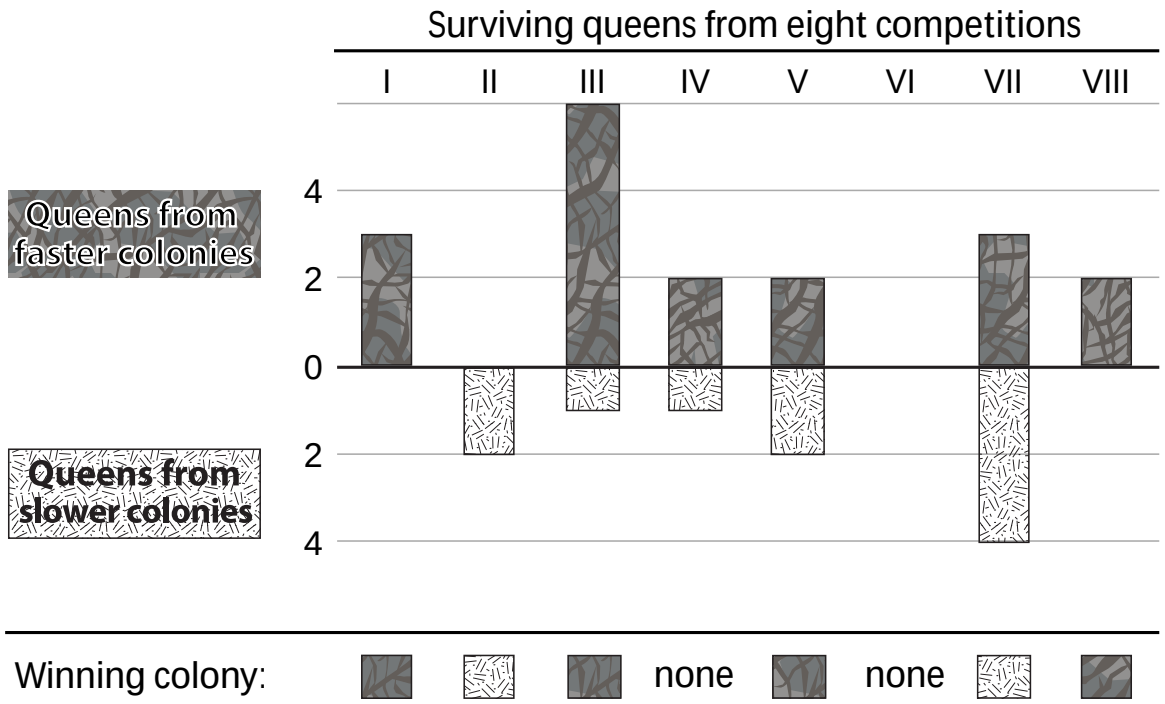


Figure 6.1: Numbers of surviving queens collected three weeks after onset of competition. No winner could be identified in competition *IV* because surviving queens had comparable ovarian development. No surviving queens were found in competition *VI*.

had comparable levels of ovarian development, making it impossible to determine a clear winner (competition *IV'* in Figure 6.1).

Gene expression differences between queens from winning and losing colonies

Nine gene annotations were significantly enriched among the expression differences between queens from the six pairs of winning and losing colonies (Table 6.1). The four most significant gene annotations we were *oxidation reduction*, *organ development*, *mitochondrial inner membrane'* and *extracellular region'* (respectively $p = 0.01$, $p = 0.019$, $p = 0.02$ and $p = 0.02$). We manually determined that genes representing these annotations were more highly expressed in queens from winning than losing colonies.

Molecular Function

GO.ID	Term	Annotated	FDR-corrected p-value
GO:0016491	oxidoreductase activity	118	0.026

Cellular Component

GO.ID	Term	Annotated	FDR-corrected p-value
GO:0005576	extracellular region	99	0.020
GO:0005743	mitochondrial inner membrane	65	0.020

Biological Process

GO.ID	Term	Annotated	FDR-corrected p-value
GO:0055114	oxidation reduction	117	0.010
GO:0048513	organ development	87	0.019
GO:0050793	regulation of developmental process	60	0.021
GO:0009653	anatomical structure morphogenesis	92	0.034
GO:0030154	cell differentiation	92	0.036
GO:0006091	generation of precursor metabolites and energy	68	0.036

Table 6.1: Gene Ontology annotations that are enriched among the genes most differently expressed between queens from winning and losing colonies (False Discovery Rate corrected p-value < 0.05)

Queens from colonies where wings are shed faster after orphaning have higher survival

Queens from colonies where queens shed their wings faster were significantly more likely to survive than queens from slower colonies (binomial GLM likelihood ratio test p = 0.009; estimated survival probabilities were respectively p=0.3 and p=0.1). The effect of competition replicate was non-significant.

Gene expression differences between queens that are faster and slower at shedding their wings after orphaning

Ten gene annotations were significantly enriched among the expression differences between queens from colonies in which queens are faster and slower at shedding their wings after orphaning (Table 6.2). The three most significant gene annotations we were *mitochondrial inner membrane*, *generation of precursor metabolites and energy* and *oxidation reduction* (respectively $p = 3.8 * 10^{-4}$, $p = 8.2 * 10^{-6}$, $p = 0.02$; see also Figure 6.2). We manually determined that genes representing these annotations were more highly expressed in queens from faster than slower colonies.

6.4 DISCUSSION

The speed at which virgin *S. invicta* queens shed their wings after orphaning could indicate how quickly they initiate reproductive development and signaling to nestmates. We thus expected queens that do this faster to produce higher quantities of olfactory signals for reproductive development. Furthermore, we

Molecular Function

GO.ID	Term	Annotated	FDR-corrected p-value
GO:0016491	oxidoreductase activity	118	0.038

Cellular Component

GO.ID	Term	Annotated	FDR-corrected p-value
GO:0005743	mitochondrial inner membrane	65	3.8E-04

Biological Process

GO.ID	Term	Annotated	FDR-corrected p-value
GO:0006091	generation of precursor metabolites and energy	68	8.2E-06
GO:0055114	oxidation reduction	117	0.020
GO:0009889	regulation of biosynthetic process	130	0.041
GO:0031326	regulation of cellular biosynthetic process	130	0.041
GO:0006139	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	292	0.041
GO:0010556	regulation of macromolecule biosynthetic process	128	0.041
GO:0031323	regulation of cellular metabolic process	154	0.041
GO:0006355	regulation of transcription, DNA-dependent	104	0.041

Table 6.2: Gene Ontology annotations that are enriched among the genes most differently expressed between queens from faster and slower colonies (False Discovery Rate corrected p-value < 0.05)

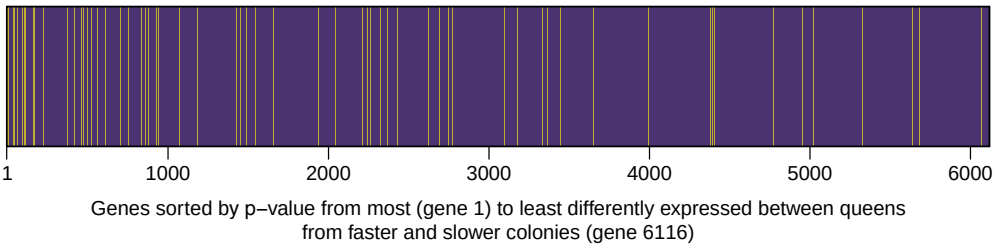


Figure 6.2: The sixty-five genes with the GO annotation *mitochondrial inner membrane* (orange) rank high according to p-values of differential expression between queens from faster and slower colonies (all other genes are in dark violet)

expected that such higher levels of reproductive signals would lead these queens to be more successful in competitions because of previous evidence that workers prefer queens that are more fertile (Fletcher & Blum 1983). To a certain extent, this was indeed the case. However, not all competitions were won by queens from colonies where queens shed their wings faster after orphaning.

Sisters of the queens that were more successful in competitions showed higher *organ development* and *cell differentiation* activities after orphaning. The genes with these annotations are likely involved in the initiation of reproductive development in orphaned virgin queens. Worker preference for these queens that show stronger organ development and cell differentiation activities may indicate two things. First, if the gene expression differences are indeed related to reproductive development, it suggests that workers preferred the more fertile queens and implies that the olfactory cues produced by queens are "honest signals" of their reproductive status. Additionally, together with the result that queens that are faster at shedding their wings do not show enrichment of annotations that indicate regulation of development, we can deduce that wing shedding and the initiation of reproductive development are processes that are regulated somewhat independently. This has been previously suggested by the finding that virgin queens from which antennae have been removed shed their wings but do not initiate reproductive development if the mother queen is still present (Vargo & Laurel 1994).

Most of the gene annotations that were overrepresented among differences between queens that were faster than slower at shedding their wings after orphaning were related to mitochondrial activity. Mitochondria play key roles in many biological processes including metabolism regulation, cell-cycle control and intra-cellular signaling (McBride *et al.* 2006). Genes related to mitochondrial activity are down-regulated within the first 24 hours of orphaning (Chapter 5), suggesting that mitochondria are unlikely to be involved in the post-orphaning changes occurring in virgin queens *per se*. It is difficult to speculate by which molecular process mitochondrial activity may be linked to the speed at which shed their wing but not the speed at which they initiate reproductive development after orphaning.

CONCLUSION

To our knowledge, this was the first study to combine behavioral experiments with large scale surveys of gene expression of whole bodies to examine a competition for reproductive dominance. We determined that subordinate *S. invicta* queens that are faster at shedding their wings after orphaning have consistently higher mitochondrial activity and are more likely to survive and than queens that are slower. Irrespective of wing shedding speed, queens that won competitions had higher mitochondrial activity as well as higher activity levels of genes related to organ development and cell differentiation, indicating that

multiple elements affect queen survival. This study provides some insight into the molecular processes behind the competition for reproductive dominance in fire ants, however much behavioral and genomic work remains to truly understand how workers choose between competing candidate queens.

SUPPORTING INFORMATION

Supporting Table 1:Colony rankings based on speed of wing shedding after orphaning.

Colony	Five individual rankings from slowest to fastest: based on the amount of time elapsed until 2 out of 10 queens shed their wings	based on the amount of time elapsed until 5 out of 10 queens shed their wings	based on the amount of time elapsed until half of the queens that shed their wings within 5 days of orphaning had shed their wings	based on the percentage of queens that had shed their wings within 5 days of orphaning	manually established ranking taking into account the cumulative percentages of queens that shed their wings after 12, 24, 36, 48, 72, and 78h	Sum of five individual rankings	Colony rank	
K	1	1	1	1	1	5	1	slowest
D	5	2	2	2	3	14	2	
A	2	4	4	5	2	17	3	
L	3	3	9	6	7	28	4	
R	10	6	3	4	5	28	5	
S	*4*	*13*	*7*	*3*	*6*	*33*	*6*	*
M	8	5	6	9	8	36	7	
F	7	9	8	11	4	39	8	
E	6	11	10	10	9	46	9	
Q	14	10	5	8	10	47	10	
H	12	7	15	7	11	52	11	
N	9	12	11	12	12	56	12	
G	11	15	13	13	13	65	13	
C	15	8	16	15	15	69	14	
P	13	16	12	16	14	71	15	
T	17	14	14	14	16	75	16	
U	16	17	17	17	17	84	17	fastest

** colony S was discarded after ranking since it contained fewer virgin queens than the other colonies*

Supporting Table 2:Pairs of colonies used for establishing competitions

Competition	Slower colony	Faster colony
I.	K	Q
II.	D	H
III.	A	N
IV.	L	G
V.	R	C
VI.	M	P
VII.	F	T
VIII.	E	U

7

OUTLOOK

by Yannick Wurm

Much of the work presented in this thesis facilitates molecular research on social behavior in ants. I also used the developed tools to examine the onset of a competition for reproductive dominance in queens of the fire ant, *Solenopsis invicta*. To begin this conclusion, I will discuss the results of the queen competition study. Subsequently, I will discuss some of the challenges and opportunities for molecular-genetic research on ants.

7.1 COMPETING QUEENS AND EXPRESSION OF GENES

A central issue in research on social life is to understand how reproductive dominance hierarchies are maintained and how individuals react to opportunities for social ascension (Solomon & French 1997). We used the fire ant as a model to examine the reaction to loss of the dominant individual on subordinate individuals at a molecular level. Upon the death of the dominant queen in the single-queen form of *S. invicta*, virgin queens shed their wings and initiate reproductive development in lieu of departing on a mating flight. Workers progressively execute almost all of them over the following weeks based on pheromonal cues produced by the young queens after orphaning (Fletcher & Blum 1981). We identified genes that are differentially expressed in the young queens after removal of the dominant individual (**Chapter 5**) and determined that the activities of mitochondrial and organ development genes are linked to the likelihood of becoming a replacement queen (**Chapter 6**). These results give rise to many additional questions about the competition between queens. Do they truly affect the likelihood of becoming a replacement queen or simply correlate with other measures? Are other elements involved that we could not detect? Are there "absolute" measurements of queen quality or do the criteria by which workers choose one queen over the other vary according to the situation?

Much more detailed studies are required to obtain a thorough overview of the molecular basis of queen choice. First, each queen lineage in our study was used only once. All possible pairwise

competitions between lineages of queens should be conducted and repeated several times. This would clarify whether there are absolute determinants of queen success, and the extent to which other, possibly stochastic elements also play roles. Second, individual discrimination within ants is largely based on odors (d'Ettorre & Moore 2008) which may go undetected when profiling gene expression of whole bodies. A comprehensive approach would thus examine competing queens both at a finer scale (perhaps individual glands or body parts) and at multiple levels (gene expression but also cuticular hydrocarbon and proteomic profiles). These multiple levels of data could be correlated to competitive success of queens and help pinpoint high confidence candidate molecules. Finally, functional followup studies will need to be performed on the candidate molecules to determine whether they indeed affect the success of competing queens. This implies ectopic activation or silencing of candidate molecules and subsequent modification of competition outcome.

7.2 SOCIOGENETIC RESEARCH ON ANTS

Conducting molecular research on non-model organisms requires custom development and fine tuning of molecular protocols and analysis tools. This can be tedious but rewarding since the newly harnessed methods can be applied to novel questions.

A successful example is the work of John Wang who had initiated the development of the molecular tools for *S. invicta* (**Chapters** 3 and 4). He used them to conduct two gene expression comparisons. On one hand he compared gene expression of workers that lack the *b* allele from single- and multiple-queen colonies. On the other hand, he compared gene expression between workers that lack and carry the *b* allele within multiple-queen colonies (See **Introduction** on page 3). Surprisingly, worker gene expression profiles are more strongly influenced by indirect effects associated with the allelic composition of the workers' colony than by the direct effect of their own genotype (Wang *et al.* 2008). This highlights the importance of social context on gene expression in individuals and links it to their behavior.

An unresolved issue is to which extent the gene encoded by the *b* allele is functionally responsible for determining social structure in fire ants (Wang *et al.* 2008). Indeed, cosegregating genes may also contribute. To elucidate the genetic architecture of the social polymorphism we have begun whole genome sequencing of the fire ant using ultra-high throughput sequencing technologies as well as construction of a genetic linkage map. It will be challenging since the technologies are young and assembly of a eukaryotic genome without a reference sequence is also still in its infancy. However we are lucky to collaborate on this project with I. Xenarios and L. Falquet of the Vital-IT High Performance computing center and L. Farinelli of Fasteris (a company that specializes in sequencing).

Following our lead, several additional ant genome projects have sprung up since the beginning of 2009. The different projects focus on species-specific questions and thus have little overlap. Importantly, molecular research on ants in general will greatly benefit from these efforts. However, the bioinformatics workload required to assemble and annotate the raw sequence data will be tremendous. Once that hurdle is overcome, many others remain before genomic reserach on ants becomes routine. Indeed, artificial selection experiments would be slow in ants since weeks to years go by between generations, controlled or artificial matings are difficult, neither transgenics nor RNA interference are routinely performed, and cell lines are non-existant.

Luckily, some questions will already benefit as raw and subsequently partially assembled or sequence become available. One of the projects for which this is the case is a collaboration between John Wang, Michael Nicolas and myself. We used the microarrays from Chapter 4 to examine gene expression in fire ant workers, queens and males at 20 timepoints between the onset of metamorphosis and eclosion (*manuscript in preparation*). We identified modules of coregulated genes that have caste-specific expression patterns. The genome projects will improve annotation of these genes and their promotor regions. That information can help identify elements responsible for coregulation of the genes and differences between castes. However, we already have preliminary results based only on current EST sequence. For instance, we discovered a gene that is expressed exclusively in queens throughout development. Surprisingly, the only publicly available sequence to which that gene shows significant similarity is one of nine genes that is 25 times more highly expressed in queens than workers of the *Lasius niger* garden ant (Gräff *et al.* 2007). The last common ancestor of *Solenopsis* and *Lasius* existed approximately 120 million years ago and ants are thought to be 140 million years old (Brady *et al.* 2006; Brady *et al.* 2009), indicating that the function of this gene may be ancestrally conserved. We are currently further characterizing the queen-specific gene in fire ants and subsequently will examine whether its expression pattern holds in other ant species.

The new availability of molecular tools for ants should improve understanding of many aspects of the function and evolution of their social lifestyles. Through which molecular mechanisms can ant queens live two orders of magnitude longer than workers (Keller & Genoud 1997)? How are collective decisions made? How do the abnormal reproductive systems identified in some ants function (See example of *Wasmannia* on page 6, others reviewed in Heinze 2008) and what is the impact of such reproductive systems on the evolution of the genome evolution? Did exclusively feeding on an obligatorily mutualistic fungus affect the digestive genes of leaf-cutter ants? Similarly, how are the genomes of obligate social parasites affected by the absence of a worker caste? Will the theoretical predictions about the evolution of caste specific genes hold? For example, Linksvayer & Wade (2009) propose that molecular evidence

for selection should be stronger on genes that are used in males and queens than on worker-specific genes because selection acts directly on queens and males but only indirectly on non-reproducing workers. Will the ant and honey-bee genomes show convergent characteristics that could be explained by social life? I expect molecular-genetic studies of ants to bring many fascinating discoveries in the years to come.

Part IV

BIBLIOGRAPHY

BIBLIOGRAPHY

Alaux, C, Y Le Conte, HA Adams, *et al.* (2009). "Regulation of brain gene expression in honey bees by brood pheromone." *Genes, Brain and Behavior* 8.3, p. 309–319. ISSN: 1601-183X (Electronic). DOI: 10.1111/j.1601-183X.2009.00480.x. See p. 5.

Alexa, A, J Rahnenführer, & T Lengauer (2006). "Improved scoring of functional groups from gene expression data by decorrelating GO graph structure". *Bioinformatics* 22.13, p. 1600–1607. See p. 72, 100.

Anderson, C & DW McShea (2001). "Individual versus social complexity, with particular reference to ant colonies". *Biol Rev Camb Philos Soc* 76.2, p. 211–37. See p. 5.

Ascunce, MS, AM Bouwma, & D Shoemaker (2009). "Characterization of 24 microsatellite markers in 11 species of fire ants in the genus *Solenopsis* (Hymenoptera: Formicidae)". *Molecular Ecology Resources*. See p. 98.

Ashburner, M, CA Ball, JA Blake, *et al.* (2000). "Gene Ontology: tool for the unification of biology." *Nature Genetics* 25.1, p. 25–29. See p. 72, 100.

Bates, D & M Maechler (online). "lme4: Linear mixed-effects models using S4 classes". URL: http://cran.r-project.org/web/packages/lme4/ (visited on //2009). See p. 99.

Ben-Shahar, Y, A Robichon, MB Sokolowski, & GE Robinson (2002). "Influence of Gene Action Across Different Time Scales on Behavior". *Science* 296.5568, p. 741–744. URL: http://www.sciencemag.org/cgi/content/abstract/296/5568/741. See p. 4, 5.

Björklund, S & CM Gustafsson (2005). "The yeast Mediator complex and its regulation". *Trends in Biochemical Sciences* 30.5, p. 240–244. See p. 79.

Bourke, AFG & NR Franks (1995). *Social Evolution in Ants*. Princeton, NJ: Princeton University Press. ISBN: 0691044279 (cl). See p. 3, 4, 70, 94.

Brady, SG, TR Schultz, BL Fisher, & PS Ward (2006). "Evaluating alternative hypotheses for the early evolution and diversification of ants". *Proceedings of the National Academy of Sciences of the USA*, p. 0605858103–. URL: http://www.pnas.org/cgi/content/abstract/0605858103v1. See p. 109.

Brady, SG, L Larkinb, & BN Danforth (2009). "The Timetree of Life". Chap. Bees, ants, and stinging wasps (Aculeata). See p. 109.

Brent, C & EL Vargo (2003). "Changes in juvenile hormone biosynthetic rates and whole body content in maturing virgin queens of *Solenopsis invicta*." *Journal of Insect Physiology* 49, p. 967–974. See p. 6, 70, 78, 94.

Buckley, KM, HE Melikian, CJ Provoda, & MT Waring (2000). "Regulation of neuronal function by protein trafficking: a role for the endosomal pathway". *Journal of Physiology* 525.(Pt 1), p. 11–19. See p. 79.

Burmeister, SS, ED Jarvis, & RD Fernald (2005). "Rapid behavioral and genomic responses to social opportunity". *PLoS Biology* 3.11, e363. DOI: 10.1371/journal.pbio.0030363. See p. 70, 94.

Burns, SN, PEA Teal, RK Vander Meer, J Nation, & JT Vogt (2002). "Identification and action of JH III from sexually mature alate females of the red imported fire ant, *Solenopsis invicta*". *Journal of Insect Physiology* 48, p. 357–365. See p. 78.

Burns, SN, RKV Meer, & PEA Teal (2007). "Juvenile Hormone III Concentrations in Female Reproductives of Solenopsis invicta Buren". *J. Entomol. Sci.* See p. 6.

Chapman, RF (1988). *The Insects: Structure and Function*. 4th Edition. Cambridge University Press. See p. 6.

Cristino, AS, FMF Nunes, CH Lobo, *et al.* (2006). "Caste development and reproduction: a genome-wide analysis of hallmarks of insect eusociality". *Insect Molecular Biology* 15.5, p. 703–714. URL: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2583.2006.00696.x. See p. 5.

Darwin, CR (1859). *On the origin of species*. John Murray. See p. 3.

Dawkins, R (1976). *The Selfish Gene*. Oxford University Press. See p. 3.

Dejean, A, PJ Solano, J Ayroles, B Corbara, & J Orivel (2005). "Insect behaviour Arboreal ants build traps to capture prey". *Nature* 434.7036, p. 973–973. URL: http://dx.doi.org/10.1038/434973a. See p. 5.

d'Ettorre, P & AJ Moore (2008). "Sociobiology of Communication: An Interdisciplinary Perspective". Ed. by P d'Ettorre & DP Hughes. Oxford University Press. Chap. Chemical communication and the coordination of social interactions in insects. See p. 108.

Dierick, HA & RJ Greenspan (2006). "Molecular analysis of flies selected for aggressive behavior". *Nature Genetics* 38.9, p. 1023–1031. URL: http://dx.doi.org/10.1038/ng1864. See p. 4.

Dietemann, V, J Pflugfelder, S Hartel, P Neumann, & RM Crewe (2006). "Social parasitism by honeybee workers (*Apis mellifera capensis* Esch.): evidence for pheromonal resistance to host queen's signals". *Behavioral Ecology and Sociobiology* 60.6 (Oct. 2006), p. 785–793. DOI: DOI10.1007/s00265-006-0222-0. See p. 70, 94.

Dobata, S, T Sasaki, H Mori, E Hasegawa, M Shimada, & K Tsuji (2009). "Cheater genotypes in the parthenogenetic ant Pristomyrmex punctatus". *Proc Biol Sci* 276.1656, p. 567–74. See p. 6.

Elango, N, BG Hunt, MAD Goodisman, & SV Yi (2009). "DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, Apis mellifera". *Proc Natl Acad Sci U S A*. DOI: 10.1073/pnas.0900301106. See p. 5.

Fletcher, DJC, D Cherix, & MS Blum (1983). "Some factors influencing dealation by virgin queen fire ants." *Insectes Sociaux* 30, p. 443–454. See p. 70, 71, 95.

Fletcher, DJC & MS Blum (1981). "Pheromonal Control of Dealation and Oogenesis in Virgin Queen Fire Ants". *Science* 212.4490, p. 73–75. See p. 70, 95, 107.

– (1983). "Regulation of Queen Number by Workers in Colonies of Social Insects". *Science* 219.4582, p. 312–314. URL: http://www.sciencemag.org/cgi/content/abstract/219/4582/312. See p. 70, 94, 95, 104.

Fournier, D, A Estoup, J Orivel, *et al.* (2005). "Clonal reproduction by males and females in the little fire ant". *Nature* 435.7046, p. 1230–1234. URL: http://dx.doi.org/10.1038/nature03705. See p. 6.

Franks, NR, AB Sendova-Franks, J Simmons, & M Mogie (1999). "Convergent evolution, superefficient teams and tempo in Old and New World army ants". *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266, p. 1697–1701. See p. 5.

Franks, NR (1985). "Reproduction, foraging efficiency and worker polymorphism in army ants." *Fortschritte der Zoologie*. See p. 5.

Frehlick, LJ, JM Eirín-López, & J Ausió (2007). "New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones". *Bioessays* 29.1, p. 49–59. See p. 79.

Gentleman, RC, VJ Carey, DM Bates, *et al.* (2004). "Bioconductor: open software development for computational biology and bioinformatics". *Genome Biology* 5, R80. URL: http://genomebiology.com/2004/5/10/R80. See p. 72, 99.

Goodisman, M, J Isoe, D Wheeler, & M Wells (2005). "Evolution of insect metamorphosis: A microarray-based study of larval and adult gene expression in the ant Camponotus festinatus". *Evolution* 59.4, p. 858–870. See p. 6.

Gotzek, D & KG Ross (2007). "Genetic Regulation of Colony Social Organization in Fire Ants: an Integrative Overview". *The Quarterly Review of Biology* 82.3, p. 201–226. URL: http://www.journals.uchicago.edu/doi/abs/10.1086/519965. See p. 6, 78.

Gräff, J, S Jemielity, JD Parker, KM Parker, & L Keller (2007). "Differential gene expression between adult queens and workers in the ant *Lasius niger*". *Molecular Ecology* 16.3, p. 675–683. See p. 6, 70, 78, 94, 109.

Grozinger, CM, NM Sharabash, CW Whitfield, & GE Robinson (2003). "Pheromone-mediated gene expression in the honey bee brain". *Proceedings of the National Academy of Sciences of the USA* 100.Suppl 2, p. 14519–14525. URL: http://www.pnas.org/content/100/suppl.2/14519.abstract. See p. 5.

Grozinger, CM, Y Fan, SER Hoover, & ML Winston (2007). "Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (*Apis mellifera*)." *Molecular Ecology* 16.22, p. 4837–4848. ISSN: 0962-1083 (Print). DOI: 10.1111/j.1365-294X.2007.03545.x. See p. 5, 73, 75.

Haider, S, B Ballester, D Smedley, J Zhang, P Rice, & A Kasprzyk (2009). "BioMart Central Portal–unified access to biological data". *Nucleic Acids Research* 37.Web Server issue, W23–27. See p. 73.

Hamilton, WD (1964). "The genetical evolution of social behaviour. I". *J Theor Biol* 7.1, p. 1–16. See p. 3, 4.

Han, YL, HM Xu, J Deng, *et al.* (2006). "Over-expression of the cellular repressor of E1A-stimulated genes inhibits the apoptosis of human vascular smooth muscle cells in vitro". *Sheng Li Xue Bao (Acta physiologica Sinica)* 58.4, p. 324–330. See p. 79.

Hedges, SB & S Kuma (2009). *The Timetree of Life*. Oxford University Press, New York. See p. 4.

Heinze, J (2008). "The demise of the standard ant (Hymenoptera: Formicidae)". *Myrmecological News*. See p. 6, 109.

Hölldobler, B & EO Wilson (1990). *The Ants*. Cambridge, Massachusetts: Belknap Press of Harvard University Press. ISBN: 0674040759 (alk. paper). See p. 5.

Holmes, MM, GJ Rosen, CL Jordan, GJ de Vries, BD Goldman, & NG Forger (2007). "Social control of brain morphology in a eusocial mammal". *Proceedings of the National Academy of Sciences of the USA* 104.25, p. 10548–10552. See p. 70, 94.

Ingram, KK, P Oefner, & DM Gordon (2005). "Task-specific expression of the foraging gene in harvester ants". *Molecular Ecology* 14.3, p. 813. See p. 7.

Jouvenaz, DP, GE Allen, WA Banks, & DP Wojcik (1977). "A Survey for Pathogens of Fire Ants, *Solenopsis* spp., in the Southeastern United States". *The Florida Entomologist* 60.4, p. 275–279. URL: http://links.jstor.org/sici?sici=0015-4040%28197712%2960%3A4%3C275%3AASFPOF%3E2.0.CO%3B2-U. See p. 71, 96.

Kauffmann, A, R Gentleman, & W Huber (2009). "arrayQualityMetrics–a Bioconductor package for quality assessment of microarray data". *Bioinformatics* 25.3, p. 415–416. See p. 72, 99.

Kearney, Toom, & Blomquist (1977). "Induction of dealation in virgin female Solenopsis invicta with juvenile hormones". *Annals of the Entomological Society of America*. See p. 6.

Keller, L & HK Reeve (1994). "Partitioning of reproduction in animal societies". English. *Trends in Ecology and Evolution* 9.3, p. 98–102. ISSN: 0169-5347. See p. 70, 94.

Keller, L (1993). *Queen Number and Sociality in Insects*. Oxford, UK: Oxford University Press. ISBN: 0198540574. See p. 70, 94.

– (1999). *Levels of selection in evolution*. Princeton, N.J.: Princeton University Press. ISBN: 0691007039 (cl : alk. paper). See p. 3.

– (2007). "Uncovering the Biodiversity of Genetic and Reproductive Systems: Time for a More Open Approach". *The American Naturalist* 169.1, p. 1–8. See p. 6.

Keller, L & M Genoud (1997). "Extraordinary lifespans in ants: a test of evolutionary theories of ageing". *Nature* 389.6654, p. 958–960. URL: http://dx.doi.org/10.1038/40130. See p. 109.

Keller, L & KG Ross (1998). "Selfish genes: a green beard in the red fire ant". *Nature* 394.6693, p. 573–575. URL: http://dx.doi.org/10.1038/29064. See p. 6, 70, 71, 78, 94, 96.

Kocher, SD, FJ Richard, DR Tarpy, & CM Grozinger (2008). "Genomic analysis of post-mating changes in the honey bee queen (*Apis mellifera*)." *BMC Genomics* 9.1, p. 232. ISSN: 1471-2164 (Electronic). DOI: 10.1186/1471-2164-9-232. See p. 5, 73, 77.

Krieger, MJB & L Keller (1997). "Polymorphism at dinucleotide microsatellite loci in fire ant Solenopsis invicta". *Molecular Ecology* 6.10, p. 997–999. DOI: doi:10.1046/j.1365-294X.1997.00264.x. See p. 98.

Krieger, MJB & KG Ross (2002). "Identification of a Major Gene Regulating Complex Social Behavior". *Science* 295.5553, p. 328–332. URL: http://www.sciencemag.org/cgi/content/abstract/295/5553/328. See p. 71, 78, 96.

Kronauer, DJC (2008). "Recent advances in army ant biology (Hymenoptera: Formicidae)". *Myrmecological news* 12, p. 51–65. See p. 5.

Langer, P, K Hogendoorn, & L Keller (2004). "Tug-of-war over reproduction in a social bee". *Nature* 428.6985, p. 844–847. URL: http://dx.doi.org/10.1038/nature02431. See p. 4.

Lawniczak, MKN & DJ Begun (2004). "A genome-wide analysis of courting and mating responses in *Drosophila melanogaster* females." *Genome* 47.5, p. 900–910. ISSN: 0831-2796 (Print). DOI: 10.1139/g04-050. See p. 73, 77.

Lenoir, A, P D'Ettorre, C Errard, & A Hefetz (2001). "Chemical ecology and social parasitism in ants". *Annu Rev Entomol* 46, p. 573–99. See p. 5.

Linksvayer, TA & MJ Wade (2009). "Genes with social effects are expected to harbor more sequence variation within and between species". *Evolution* 63.7, p. 1685–96. See p. 109.

Lodish, H, A Berk, SL Zipursky, P Matsudaira, D Baltimore, & J Darnell (2000). *Molecular Cell Biology*. W.H Freeman & Company, New York. See p. 79.

Lomberk, G, L Wallrath, & R Urrutia (2006). "The Heterochromatin Protein 1 family". *Genome Biology* 7.7, p. 228. See p. 79.

Lucas, C & MB Sokolowski (2009). "Molecular basis for changes in behavioral state in ant social behaviors". *Proceedings of the National Academy of Sciences of the United States of America* 106.15 (Apr. 2009), p. 6351–6356. DOI: DOI10.1073/pnas.0809463106. See p. 7.

Manoli, DS, M Foss, A Villella, BJ Taylor, JC Hall, & BS Baker (2005). "Male-specific fruitless specifies the neural substrates of Drosophila courtship behaviour". *Nature* 436.7049, p. 395–400. URL: http://dx.doi.org/10.1038/nature03859. See p. 4.

Maynard Smith, J & E Szathmáry (1995). *The major transitions in evolution*. Oxford: W.H. Freeman Spektrum. ISBN: 0716745259. See p. 3.

McBride, HM, M Neuspiel, & S Wasiak (2006). "Mitochondria: more than just a powerhouse". *Curr Biol* 16.14, R551–60. See p. 104.

McGraw, LA, G Gibson, AG Clark, & MF Wolfner (2004). "Genes regulated by mating, sperm, or seminal proteins in mated female *Drosophila melanogaster*." *Current Biology* 14.16, p. 1509–1514. ISSN: 0960-9822 (Print). DOI: 10.1016/j.cub.2004.08.028. See p. 73, 77.

Montellano, PRO de (2005). *Cytochrome P450: Structure, Mechanism, and Biochemistry*. Kluwer Academic/Plenum Publisher, New York. See p. 77.

Neumann, P, HR Hepburn, & SE Radloff (2000). "Modes of worker reproduction, reproductive dominance and brood cell construction in queenless honeybee (*Apis mellifera* L.) colonies". *Apidologie* 31.4, p. 479–486. See p. 70, 94.

Nielsen, MG (2000). "Distribution of the ant (Hymenoptera:Formicidae) fauna in the canopy of the mangrove tree Sonneratia alba J.Smith in northern Australia". *Australian Journal of Entomology* 39, p. 275–279. See p. 5.

Osborne, KA, A Robichon, E Burgess, *et al.* (1997). "Natural Behavior Polymorphism Due to a cGMP-Dependent Protein Kinase of Drosophila". *Science* 277.5327, p. 834–836. URL: http://www.sciencemag.org/cgi/content/abstract/277/5327/834. See p. 4.

Passera, L & S Aron (2005). *Les fourmis. Comportement, Organisation sociale et évolution*. Presses scientifique du CNRC, Conseil National de Recherches. See p. 4.

Patel, A, MK Fondrk, O Kaftanoglu, *et al.* (2007). "The Making of a Queen: TOR Pathway Is a Key Player in Diphenic Caste Development". *PLoS ONE* 2.6, e509. DOI: 10.1371/journal.pone.0000509. URL: http://dx.doi.org/10.1371%2Fjournal.pone.0000509. See p. 5.

Pearcy, M, S Aron, C Doums, & L Keller (2004). "Conditional use of sex and parthenogenesis for worker and queen production in ants." *Science* 306.5702, p. 1780–3. See p. 6.

Pelosi, P, M Calvello, & L Ban (2005). "Diversity of odorant-binding proteins and chemosensory proteins in insects". *Chemical Senses* 30.Suppl 1, p. i291–i292. See p. 78.

Pereboom, JJM, WC Jordan, S Sumner, RL Hammond, & AFG Bourke (2005). "Differential gene expression in queen–worker caste determination in bumble-bees". *Proceedings of the Royal Society B: Biological Sciences* 272.1568 (06/07/2005), p. 1145–1152. URL: http://dx.doi.org/10.1098/rspb.2005.3060. See p. 70, 94.

Queller, DC & JE Strassmann (1988). "Kin Selection and Social Insects". *BioScience*. See p. 4.

Remm, M, CEV Storm, & ELL Sonnhammer (2001). "Automatic clustering of orthologs and in-paralogs from pairwise species comparisons". *Journal of Molecular Biology* 314.5 (Dec. 2001), p. 1041–1052. URL: http://www.sciencedirect.com/science/article/B6WK7-45M7TB0-5T/2/941f5abac97e03f80ad7bccd5f2fbee3. See p. 73.

Robinson, GE (1992). "Regulation of division of labor in insect societies". *Annu Rev Entomol* 37, p. 637–65. See p. 5.

Robinson, GE & EL Vargo (1997). "Juvenile hormone in adult eusocial Hymenoptera: gonadotropin and behavioral pacemaker." *Archives of Insect Biochemistry and Physiology* 35.4, p. 559–583. See p. 78.

Rogers, DW, MMA Whitten, J Thailayil, J Soichot, EA Levashina, & F Catteruccia (2008). "Molecular and cellular components of the mating machinery in *Anopheles gambiae* females." *Proceedings of the National Academy of Sciences of the USA* 105.49, p. 19390–19395. ISSN: 1091-6490 (Electronic). DOI: 10.1073/pnas.0809723105. See p. 73, 77.

Roseler, P (1991). "Social and reproductive dominance among ants". *Naturwissenschaften* 78.3 (Mar. 1991), p. 114–120. See p. 70, 94.

Roseler, P, I Roseler, A Strambi, & R Augier (1984). "Influence of insect hormones on the establishment of dominance hierarchies among foundresses of the paper wasp, *Polistes gallicus*". *Behavioral Ecology and Sociobiology* 15.2, p. 133–142. See p. 70, 94.

Ross, KG & L Keller (1998). "Genetic control of social organization in an ant." *Proceedings of the National Academy of Sciences of the USA* 95.24, p. 14232–14237. ISSN: 0027-8424 (Print). See p. 6, 78.

– (2002). "Experimental conversion of colony social organization by manipulation of worker genotype composition in fire ants (*Solenopsis invicta*)". *Behavioral Ecology and Sociobiology* 51.3,

p. 287–295. URL: http://www.springerlink.com/openurl.asp?genre=article\&id=doi:10.1007/s00265-001-0431-5. See p. 78.

Ross, KG (1997). "Multilocus evolution in fire ants: effects of selection, gene flow and recombination". *Genetics* 145.4, p. 961–974. See p. 71, 96.

Schultz, TR & SG Brady (2008). "Major evolutionary transitions in ant agriculture". *Proc Natl Acad Sci U S A* 105.14, p. 5435–40. See p. 5.

Schwander, T & L Keller (2008). "Genetic compatibility affects queen and worker caste determination". *Science* 322.5901, p. 552. See p. 6.

Shoemaker, DD, C DeHeer, MJB Krieger, & KG Ross (2006). "Population genetics of the invasive fire ant *Solenopsis invicta* in the U.S.A". *Annals of the Entomological Society of America* 99.6, p. 1213–1233. See p. 71, 96.

Smyth, GK (2004). "Linear models and empirical bayes methods for assessing differential expression in microarray experiments". *Statistical Applications in Genetics and Molecular Biology* 3.1, Article 3. URL: http://www.bepress.com/sagmb/vol3/iss1/art3. See p. 72, 99.

Sokolowski, MB (1980). "Foraging strategies of *Drosophila melanogaster*: A chromosomal analysis". *Behavior Genetics* 10.3 (05/01/1980), p. 291–302. See p. 7.

Solomon, NG & JA French (1997). *Cooperative Breeding in Mammals*. Cambridge, U.K.: Cambridge University Press. ISBN: 0521454913 (hc). See p. 70, 94, 107.

Strassmann, JE (2006). "Social evolution: early production of deadly males by competing queens". *Curr Biol* 16.24, R1023–5. See p. 4.

Subramanian, A, P Tamayo, VK Mootha, *et al.* (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles". *Proc Natl Acad Sci U S A* 102.43, p. 15545–50. See p. 100.

The Honey Bee Genome Sequencing Consortium (2006). "Insights into social insects from the genome of the honeybee *Apis mellifera*". *Nature* 443.7114, p. 931–949. URL: http://dx.doi.org/10.1038/nature05260. See p. 4, 70, 94.

The R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. See p. 72, 99.

Tian, H, SB Vinson, & C Coates (2004). "Differential gene expression between alate and dealate queens in the red imported fire ant, *Solenopsis invicta* Buren (Hymenoptera: Formicidae)." *Insect Biochemistry and Molecular Biology* 34.9, p. 937–949. See p. 6, 78, 79.

Toth, AL, K Varala, TC Newman, *et al.* (2007). "Wasp Gene Expression Supports an Evolutionary Link Between Maternal Behavior and Eusociality". *Science* 318.5849, p. 441–444. URL: http://www.sciencemag.org/cgi/content/abstract/318/5849/441. See p. 70, 95.

Troisi, S & L Riddiford (1974). "Juvenile hormone effects on metamorphosis and reproduction of the fire ant, Solenopsis invicta". *Environmental Entomology*. See p. 6.

Tschinkel, WR (2006). *The Fire Ants*. Cambridge, Massachusetts: The Belknap Press of Harvard University Press. ISBN: 0674022076. See p. 71, 78, 95.

Vander Meer, RK, BM Glancey, CS Lofgren, A Glover, JH Tumlinson, & J Rocca (1980). "The poison gland of the red imported fire ant queens: source of a pheromone attractant". *Annals of the Entomological Society of America* 73, p. 609–612. See p. 70, 94.

Vargo, EL & CD Hulsey (2000). "Multiple glandular origins of queen pheromones in the fire ant *Solenopsis invicta*". *Journal of Insect Physiology* 46.8, p. 1151–1159. See p. 70, 94.

Vargo, EL & M Laurel (1994). "Studies on the mode of action of a queen primer pheromone of the fire ant *Solenopsis invicta*". *Journal of Insect Physiology* 40.7, p. 601–610. See p. 6, 70, 77–79, 94, 95, 104.

Vargo, EL & SD Porter (1993). "Reproduction by virgin fire queen ants in queenless colonies: comparative study of three taxa (*S. richteri*, hybrid *S. invicta/richteri*, *S. geminata*) (Hymenoptera:Formicidae)". *Insectes Sociaux* 40, p. 283–293. See p. 70, 95.

Vargo, EL (1999). "Reproductive development and ontogeny of queen pheromone production in the fire ant *Solenopsis invicta*". *Physiological Entomology* 24, p. 370–376. See p. 70, 71, 79, 94, 95.

Voigt, C, S Leitner, & M Gahr (2007). "Socially induced brain differentiation in a cooperatively breeding songbird". *Proceedings of the Royal Society B: Biological Sciences* 274.1626, p. 2645–2651. See p. 70, 94.

Wang, J, S Jemielity, P Uva, Y Wurm, J Gräff, & L Keller (2007). "An annotated cDNA library and microarray for large-scale gene-expression studies in the ant *Solenopsis invicta*". *Genome Biology* 8.1, R9. ISSN: 1465-6906. DOI: 10.1186/gb-2007-8-1-r9. URL: http://genomebiology.com/2007/8/1/R9. See p. 70, 72, 94, 99.

Wang, J, KG Ross, & L Keller (2008). "Genome-Wide Expression Patterns and the Genetic Architecture of a Fundamental Social Trait". *PLoS Genetics* 4.7, e1000127. DOI: 10.1371/journal.pgen.1000127. URL: http://dx.doi.org/10.1371%2Fjournal.pgen.1000127. See p. 70, 72, 95, 99, 108.

Weil, T, J Korb, & M Rehli (2009). "Comparison of queen-specific gene expression in related lower termite species". *Molecular Biology and Evolution* 26, p. 1841–1850. DOI: 10.1093/molbev/msp095. See p. 70, 94.

West, SA, AS Griffin, & A Gardner (2007). "Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection". *J Evol Biol* 20.2, p. 415–32. See p. 3.

Wheeler, DE, N Buck, & JD Evans (2006). "Expression of insulin pathway genes during the period of caste determination in the honey bee, Apis mellifera". *Insect Molecular Biology* 15.5, p. 597–602. URL: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-2583.2006.00681.x. See p. 5.

White, KP, SA Rifkin, P Hurban, & DS Hogness (1999). "Microarray analysis of Drosophila development during metamorphosis." *Science* 286.5447, p. 2179–84. See p. 4.

White, SA, T Nguyen, & RD Fernald (2002). "Social regulation of gonadotropin-releasing hormone". *Journal of Experimental Biology* 205.Pt 17, p. 2567–2581. See p. 70, 94.

Whitfield, CW, MR Band, MF Bonaldo, *et al.* (2002). "Annotated Expressed Sequence Tags and cDNA Microarrays for Studies of Brain and Behavior in the Honey Bee". *Genome Research*. See p. 4.

Whitfield, CW, AM Cziko, & GE Robinson (2003). "Gene Expression Profiles in the Brain Predict Behavior in Individual Honey Bees". *Science* 302.5643, p. 296–299. URL: http://www.sciencemag.org/cgi/content/abstract/302/5643/296. See p. 4.

Whitfield, CW, Y Ben-Shahar, C Brillet, *et al.* (2006). "Inaugural Article: Genomic dissection of behavioral maturation in the honey bee". *Proceedings of the National Academy of Sciences of the USA* 103.44, p. 16068–16075. URL: http://www.pnas.org/cgi/content/abstract/103/44/16068. See p. 4, 5.

Wiegmann, BM, Jw Kim, & MD Trautwein (2009). "*Holometabolous insects (Holometabola). In:* The Timetree of Life". Ed. by SB Hedges & S Kumar. Oxford University Press, p. 260–263. See p. 79.

Wilson, EO (1971). *The Insect Societies*. Cambridge, Massachusetts: Belknap Press of Harvard University Press. ISBN: 0674454901. See p. 3–5, 70, 94.

Wurm, Y, P Uva, F Ricci, *et al.* (2009). "Fourmidable: a database for ant genomics." *BMC Genomics* 10, p. 5. ISSN: 1471-2164 (Electronic). DOI: 10.1186/1471-2164-10-5. See p. 70, 72, 94, 99, 100.