A Single SNP Turns a Social Honey Bee (*Apis mellifera*) Worker into a Selfish Parasite

Denise Aumer,*[,1] Eckart Stolle,*[,†,1] Michael Allsopp,[2] Fiona Mumoki,[3] Christian W.W. Pirk,[3] and Robin F.A. Moritz[†,1,3,4]

[1]Institute of Biology, Martin-Luther-Universität Halle-Wittenberg, Halle, Saale, Germany

[2]Honey Bee Research Section, ARC Plant Protection Research Institute, Stellenbosch, South Africa

[3]Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa

[4]Department of Sericulture and Apiculture, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: denise.aumer@zoologie.uni-halle.de; eckart.stolle@zoologie.uni-halle.de.

Associate editor: John True

Abstract

The evolution of altruism in complex insect societies is arguably one of the major transitions in evolution and inclusive fitness theory plausibly explains why this is an evolutionary stable strategy. Yet, workers of the South African Cape honey bee (*Apis mellifera capensis*) can reverse to selfish behavior by becoming social parasites and parthenogenetically producing female offspring (thelytoky). Using a joint mapping and population genomics approach, in combination with a time-course transcript abundance dynamics analysis, we show that a single nucleotide polymorphism at the mapped thelytoky locus (*Th*) is associated with the iconic thelytokous phenotype. *Th* forms a linkage group with the ecdysis-triggering hormone receptor (*Ethr*) within a nonrecombining region under strong selection in the genome. A balanced detrimental allele system plausibly explains why the trait is specific to *A. m. capensis* and cannot easily establish itself into genomes of other honey bee subspecies.

Key words: **social evolution, inclusive fitness, balancing selection, social parasitism, worker reproduction, thelytoky.**

Article

Introduction

The evolution of insect societies is one of the most remarkable evolutionary transitions regarding the complexity of biotic structures. Inclusive fitness theory (Hamilton 1964) convincingly explains why the altruistic behavior of female workers that sacrifice their individual fitness in favor of a closely related highly fecund queen is an evolutionary stable strategy. However, there are various exceptions in the social Hymenoptera, where workers have regained reproductive control and in some cases even evolved into social parasites. The best studied example of worker social parasitism is that of the Cape honey bee, *Apis mellifera capensis* (Neumann and Moritz 2002). Typically, reproduction in honey bee colonies is confined to the queen and under strict pheromonal control, especially that of the queen (Slessor et al. 1988; Plettner et al. 1993). However, in the case of a queen loss, workers can activate their ovaries and produce queen-like pheromones, becoming "false queens" or "pseudoqueens" (Sakagami 1958; Crewe and Velthuis 1980). In general, these pseudoqueens produce male offspring, as these workers are not mated and hence lay unfertilized haploid eggs that develop into drones (arrhenotoky; Ruttner and Hesse 1981). This is different in *A. m. capensis* workers, which produce diploid female offspring (thelytoky; de Villiers 1883; Onions 1914) via central fusion of meiotic products (automixis; Verma and Ruttner 1983). In addition, *A. m. capensis* pseudoqueens can rapidly activate their ovaries (usually within 8 days [Ruttner and Hesse 1981; Hepburn et al. 1991] and occasionally even in the presence of a queen [Pirk et al. 2002]) and can produce much higher quantities of queen-like pheromones than workers of other subspecies (Hemmling et al. 1979; Crewe and Velthuis 1980; Okosun et al. 2017). The combination of thelytokous reproduction, swift ovary activation, and queen-like pheromone secretions (thelytoky syndrome [Lattorff and Moritz 2013]) of *A. m. capensis* workers allows them to exploit an additional life history trajectory as social parasites. After entering a foreign colony, they are able to establish themselves as reproductive dominant pseudoqueens, to the detriment of the resident queen (Neumann and Moritz 2002). These social parasites are particularly successful and commonly observed in the NE of South Africa, where a single worker lineage (Baudry et al. 2004) infests colonies of the adjacent subspecies *A. m. scutellata*, causing substantial colony losses to beekeepers since decades (Allsopp 1992; Pirk et al. 2014). In addition, thelytokous worker reproduction enables queenless colonies to rapidly requeen themselves from diploid laying worker's offspring (Holmes et al. 2010), increasing the chance of colony survival, facilitated by a genetically related individual. This ability has been suggested as reason for the evolution of thelytoky in *A. m. capensis* as the

endemic region of this subspecies is characterized by suddenly changing (Tribe 1983) and very windy weather, which might make it difficult for queens to return to their colonies after mating flights (Moritz 1986). Given these selective advantages and fitness benefits, it has been difficult to explain why thelytoky did not spread further into other honey bee populations (Moritz 1986; Greeff 1996). This was particularly puzzling because the syndrome showed Mendelian segregation in experimental crosses, suggesting a single locus control (Ruttner 1988; Lattorff et al. 2005, 2007; Aumer et al. 2017). Previous studies using microsatellite quantitative trait locus mapping (Lattorff et al. 2007) and RNAi knockdown experiments (Jarosch et al. 2011) suggested that differential splicing of the transcription factor *gemini* is controlling thelytokous reproduction. However, subsequent studies (Chapman et al. 2015; Aumer et al. 2017) showed that even though *gemini* is a key regulating gene for worker reproduction in general, it was not the ultimate gene switch controlling the mode of parthenogenesis and thus the entire thelytoky syndrome.

To determine the actual genetic locus controlling thelytoky, we used an *A. m. capensis* mapping population (Aumer et al. 2017), consisting of workers produced by a single queen that had naturally mated in the wild with a large number of drones. This allowed to study both, 1) the frequency of the parental male alleles for a population-wide analysis and 2) the segregation of the heterozygous (Aumer et al. 2017) maternal alleles in a classical genetic mapping study by comparing thelytokous and arrhenotokous workers, providing a multifaceted basis for the discovery and analysis of the functional loci involved in this iconic reproductive trait (Schaid et al. 2018).

Results

The Thelytoky-Associated Locus

To identify the thelytoky-associated locus, we analyzed 21 thelytokous and 21 arrhenotokous (from 13 individual patrilines each) robustly phenotyped *A. m. capensis* workers of the mapping population using a data set of 7,238,467 biallelic single nucleotide polymorphisms (SNPs) in a genome-wide association study. A single genomic region on chromosome 1 (Group1.23: 480,000–509,450 bp) that harbors two genes, *mycosubtilin synthase subunit C* (*mycC*, LOC409109, the ortholog of *Drosophila melanogaster ebony*) and the 3′ part of the uncharacterized gene LOC409096, showed highly elevated genetic differentiation between the two phenotypic groups (fig. 1*a* and supplementary fig. 1*a*, Supplementary Material online). Conversely, the adjacent region (Group1.23: 509,450–570,000 bp) with the 5′ part of LOC409096 and the gene *ecdysis-triggering hormone receptor* (*Ethr*, LOC724495) almost completely lacked any genetic differentiation (supplementary fig. 1*a*, Supplementary Material online) as well as recombination (fig. 2). Both thelytokous and arrhenotokous workers were nearly entirely homozygous within the stretch of 60,550 bp (median of 39 [26–73] heterozygous SNPs per individual, fig. 1*b* and supplementary figs. 1*b* and 2, Supplementary Material online). Thus, the gene *Ethr* is represented by a single functional allele in *A. m. capensis* (one polymorphic nonsynonymous SNP at the 3′ end caused marginal putative changes of the protein, supplementary fig. 3, Supplementary Material online and supplementary table 1, Supplementary Material online). In contrast, the gene *mycC* showed increased levels of heterozygosity in thelytokous and arrhenotokous workers (median of 216 [185–236] heterozygous SNPs in thelytokous workers and median of 53 [1–95] heterozygous SNPs in arrhenotokous workers within 9,576 bp gene sequence, fig. 1*b* and supplementary figs. 1*b* and 2, Supplementary Material online). With increasing distance to the almost entirely homozygous region around *Ethr*, heterozygosity levels within *mycC* decreased in thelytokous workers, but increased in arrhenotokous workers toward the genomic average, consistent with a reduction in linkage disequilibrium due to recombination (figs. 1*b* and 2). Accordingly, considerable allelic variance was found within the mapping population (17 different alleles were identified based on 27 polymorphic nonsynonymous SNPs, supplementary table 1, Supplementary Material online). The lack of allelic covariance with the worker reproductive phenotype suggests that neither *Ethr* nor *mycC* control the mode of parthenogenesis. However, the alleles of LOC409096 showed consistent differentiation patterns between the two phenotypic groups. The arrhenotokous workers were almost completely monomorphic for the entire gene sequence (median of 5 [1–14] heterozygous SNPs per individual within 14,929 bp, fig. 1*c* and supplementary fig. 2, Supplementary Material online) and homozygous for a single functional allele (coding sequence as well as associated 3′ and 5′ untranslated regions [UTRs], supplementary table 1, Supplementary Material online). The thelytokous workers were similarly monomorphic in the larger 5′ part of the gene (median of 8 [4–13] heterozygous SNPs per individual within 13,435 bp) but were highly heterozygous in the 857 bp section at the 3′ end (median of 15 [12–17] heterozygous SNPs per individual, fig. 1*c* and supplementary fig. 2, Supplementary Material online), including a small part of exon 7 and the complete exon 8. As a consequence, all thelytokous workers were heterozygous with two functional alleles: one identical to the allele found in all arrhenotokous workers, the other allele specific to heterozygous thelytokous workers (supplementary table 1, Supplementary Material online). Thus, the thelytoky-specific allele appears to be dominant rather than recessive, as previously inferred in several studies (Ruttner 1988; Lattorff et al. 2005, 2007; Aumer et al. 2017). An independent data set of workers from the socially parasitic lineage (thelytokous *A. m. capensis*) in NE South Africa ($n = 4$) confirmed the consistent pattern of heterozygosity at this locus (supplementary figs. 1*c*, 1*d*, and 2, Supplementary Material online and supplementary table 1, Supplementary Material online). Due to the unambiguous association of a single allele with the mode of parthenogenesis, we suggest to term LOC409096 *Thelytoky* (*Th*), as the major regulator of thelytokous reproduction. Structural and conserved domain analyses of the *Th* amino acid sequence revealed that *Th* encodes a receptor protein with a transmembrane helix (AA430–480) and a signal peptide (AA1–21) at the extracellular N-terminus,

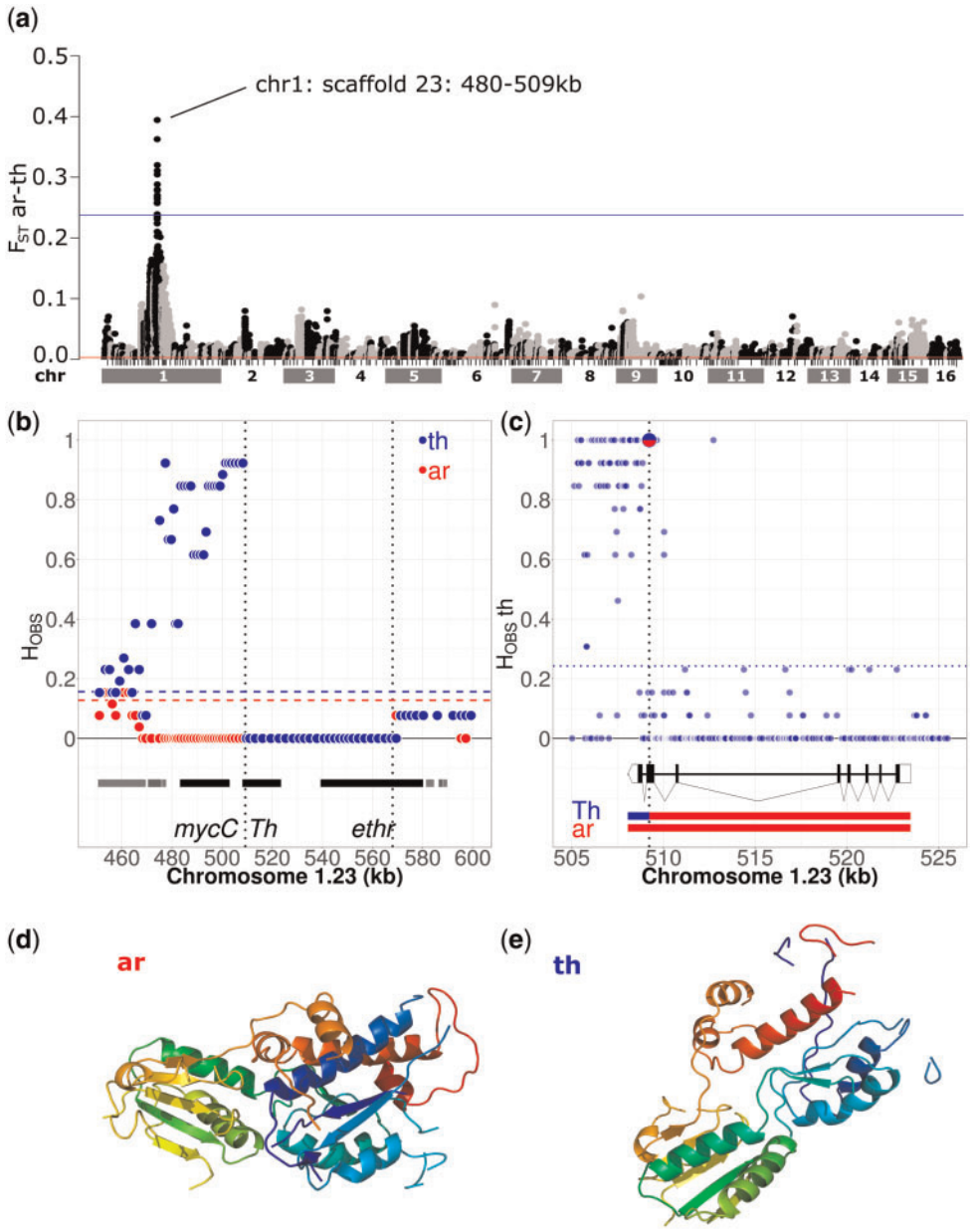

FIG. 1. The thelytoky locus in *Apis mellifera capensis*. (*a*) Genetic differentiation (F_{ST}) across the whole genome (chromosome 1–16) between the thelytokous (th) and arrhenotokous (ar) workers, shown as mean per 100 SNP sliding window (50 SNP steps, blue solid line: 99.99th percentile). (*b*) Observed heterozygosity (H_{obs} per individual) of arrhenotokous (red) and thelytokous (blue) workers (robustly phenotyped $n = 13$ per group) within the genomic region of *mycC*, *Th*, and *Ethr*, shown as mean per 100 SNP sliding window with 50 SNP steps (dashed lines: genome-wide means of both phenotypic groups, dotted vertical lines: borders of the heterozygous and homozygous regions). (*c*) Observed heterozygosity (H_{obs} per individual) of the heterozygous thelytokous (blue) workers (robustly phenotyped $n = 13$) within the genomic region of *Th* shown per SNP (dashed line: genome-wide mean). The heterozygous nonsynonymous substitution at 509,225 bp leading to the two distinct alleles (Th and ar) is displayed larger and shown in bicolor (red and blue). (*d* and *e*) Modeled tertiary protein structures for both *Th*-alleles: arrhenotoky allele (*d*) and thelytoky allele (*e*) in rainbow colors from the N-terminus (blue) to the C-terminus (red).

indicating that it is linked to the secretory pathway (Blobel and Dobberstein 1975).

The pattern of divergence and heterozygosity at the thelytoky locus is consistent with the absence of recombination between the two queen alleles within the mapping population ($n = 70$, first detected recombination event upstream of *mycC* at about 471,500 bp, first detected recombination event downstream of *Ethr* at about 593,500 bp). Since recombination suppression is often associated with structural variants such as inversions (Kirkpatrick 2010), we examined the short read sequencing data from the mapping population as well as additional Oxford Nanopore long read sequence data of two thelytokous pseudoqueens of the socially parasitic lineage for the presence of inversions and other structural variants. No evidence for an inversion was found in this region of the genome (supplementary figs. 4–10, Supplementary

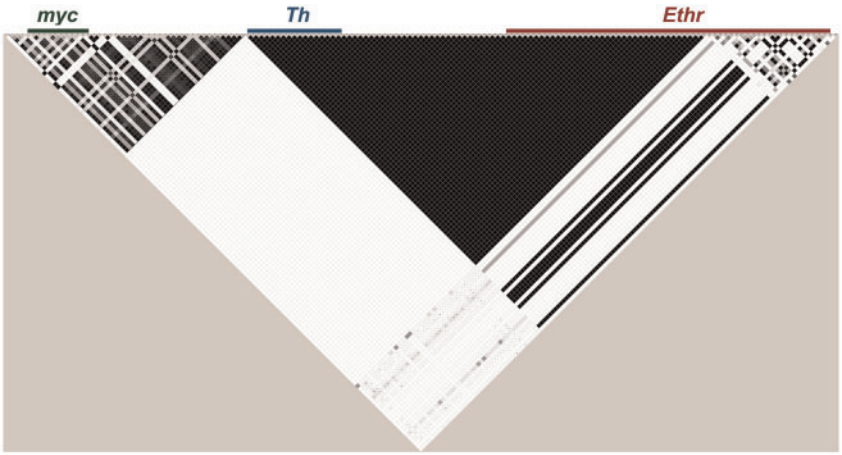

Fig. 2. The linkage disequilibrium within the genomic locus harboring *mycC*, *Th*, and *Ethr* (480,000–580,000 bp). The linkage disequilibrium was assessed using the squared Pearson coefficient of correlation (r^2) measurement (color scheme: black indicates $r^2 = 1$ [complete linkage disequilibrium/nonrandom association of two loci]; shades of gray indicate $0 < r^2 < 1$ [the darker, the bigger r^2]; white indicates $r^2 = 0$ [complete linkage equilibrium/random association of two loci]). The analysis was done with SNPs at 500-bp intervals. Each square represents the comparison of two SNPs. The location of the three genes is indicated by bars above the figure (*mycC* = green, *Th* = blue, and *Ethr* = red).

Material online). The strongly reduced recombination rate that is typical for thelytokous *A. m. capensis* workers (Baudry et al. 2004) as well as the close proximity of the thelytoky locus to the centromere, where recombination is often reduced (Baudry et al. 2004; Stapley et al. 2017), might contribute to the maintenance of the observed heterozygosity patterns at this locus.

A Single SNP Is Associated with Thelytokous Worker Reproduction in *A. m. capensis*

Within the mapping population, 63 polymorphic SNPs were identified in the coding sequence of *Th*, including three nonsynonymous SNPs, predicted to affect the amino acid sequence (supplementary table 1, Supplementary Material online). Two of these SNPs corresponded to subfamily-specific polymorphisms which were rarely observed in both phenotypic groups, leading to changes within the same amino acid group and thus, small structural changes at the protein level. Only a single nonsynonymous SNP within the 3′ part of *Th* (position 509,225 bp in Group1.23) was consistently associated with the mode of parthenogenesis. This SNP was homozygous in all arrhenotokous workers (genotype G/G), but heterozygous in all thelytokous workers (G/A, supplementary table 1, Supplementary Material online). The independent samples of thelytokous *A. m. capensis* social parasites ($n = 4$) carried the identical heterozygous SNP genotype (genotype G/A) at this position (supplementary table 1, Supplementary Material online). This substitution causes a change from the polar amino acid threonine to the nonpolar amino acid isoleucine in the protein sequence (p.Thr400Ile), leading to substantial structural modifications and likely functional consequences. Various protein prediction models revealed profound differences between the tertiary structures of the arrhenotoky (fig. 1*d*) and the thelytoky allele (fig. 1*e*), including changes of the positions and the number of the predicted protein binding sites.

The Thelytoky Locus Is Under Strong Selection in *A. m. capensis*

While the genomic regions flanking the *Th–Ethr* linkage group (60,000 kb up- and downstream) did not deviate from the genome-wide average for nucleotide diversity or heterozygosity, the linkage group of the 5′ part of *Th* and the neighboring gene *Ethr* was characterized by extremely low levels of nucleotide diversity and heterozygosity over a region of approximately 60,550 bp (fig. 1*b* and supplementary fig. 1*b–e*, Supplementary Material online). All individuals of the mapping population were mostly homozygous in this region, with the majority of SNPs (77.33%) being the derived allele (supplementary fig. 2, Supplementary Material online). These patterns were mirrored in the thelytokous socially parasitic *A. m. capensis* lineage of NE South Africa ($n = 4$). Such a drastic reduction of genetic differentiation together with the excessive linkage disequilibrium (fig. 2) typically represents exceptionally strong selection (Jackson et al. 2015). Indeed, a McDonald–Kreitman test confirmed that *Ethr* is under positive selection, consistent with a strong selective sweep with sharply reduced genetic diversity, characterized by an overabundance of fixed synonymous and nonsynonymous substitutions (supplementary table 2, Supplementary Material online). In contrast, the gene *Th* shows signatures of negative selection overall, but within thelytokous workers, the 3′ part of *Th* is under balancing selection with an excess of polymorphic sites (14 of 15 synonymous and nonsynonymous SNPs were polymorphic within 6% of the gene [626 bp]) and few nonsynonymous substitutions (2 within 6% of the gene). This and the observed elevated levels of heterozygosity (figs. 1 and 2 and supplementary fig. 2, Supplementary Material online) are indications for a locus with heterozygote advantage (overdominance). The adjacent gene *mycC* showed signatures of positive selection in arrhenotokous workers, but negative selection in thelytokous workers (supplementary

table 2, Supplementary Material online) due to close linkage (and therefore increased heterozygosity) with the Th_{Th} allele.

Population Genetics of *Th* in the Natural *A. m. capensis* Population

All 70 analyzed workers of the mapping population were offspring of a single heterozygous queen (Aumer et al. 2017) (Th_{Th}/Th_{ar}) that had mated with a large number of drones in the wild. Thus, the parental male genomes reflect a random sample of the natural *A. m. capensis* population from the Cape of Good Hope Nature Reserve, allowing us to estimate population-wide allele frequencies at *Th*. In total, we inferred the genotypes of 36 parental drones (based on the results of Aumer et al. [2017] and 100 SNPs). While the arrhenotoky allele Th_{ar} was carried by 35 drones, the thelytoky allele Th_{Th} was only found in a single drone. Considering both queen alleles (Th_{Th}/Th_{ar}), the overall allele frequencies were $p_{ar} = 0.947$ and $p_{Th} = 0.053$. Given that we never observed homozygous Th_{Th}/Th_{Th} workers (neither among the workers from the mapping population nor in the social parasitic workers from NE South Africa) and taking into account the balancing selection on the 3′ end of *Th*, we concluded that homozygosity of the Th_{Th} allele is likely associated with substantial fitness disadvantages, due to either the failure of developing into adults or infertility. The observed low frequency of Th_{Th} in parental males did, however, not reflect lethal effects on haploid drones. Nine out of 19 drone offspring of the heterozygous mapping population's queen (Th_{Th}/Th_{ar}) carried the Th_{Th} and 10 the alternative Th_{ar} allele. As this perfectly matches a 1:1 segregation, we excluded that the hemizygous Th_{Th} allele has any lethal effects on drones. This leaves potential fitness-reducing effects to the female sex. Assuming the most extreme, a fitness of zero in Th_{Th}/Th_{Th} homozygous females ($w_{Th/Th} = 0$) and heterozygous Th_{Th}/Th_{ar} having a higher relative fitness than homozygous Th_{ar}/Th_{ar} females, we estimated a fitness difference of 5.6% between the two genotypes ($w_{Th/ar} = 1$, $w_{ar/ar} = 0.944$), considering the equilibrium frequency $p_{ar} = 0.947$. Therefore, even in the most extreme case, we predict no large fitness differences between Th_{ar}/Th_{ar} and Th_{Th}/Th_{ar} queens that would be detected at the colony level.

Comparison to Other Subspecies

To determine whether other honey bee subspecies show signs of selection in the *Th–Ethr* region similar to *A. m. capensis*, we screened 53 previously published *A. mellifera* genomes (24 *A. m. scutellata* workers from Kenya [Cridland et al. 2017; Wallberg et al. 2017], 10 *A. m. scutellata* workers from South Africa, 9 *A. m. carnica* workers from Europe, and 10 *A. m. yemenitica* workers from the Arabic Peninsula [Harpur et al. 2014]). We found that in all subspecies, the levels of nucleotide diversity and heterozygosity at the *Th–Ethr* locus did not deviate from the genome-wide averages (supplementary fig. 1*f–m*, Supplementary Material online) and no signs for extreme selection were detected. Furthermore, there was substantial variability within the *Th–Ethr* locus among the analyzed workers. A single South African *A. m. scutellata* worker was homozygous for the functional alleles of *Th* and *Ethr* found in arrhenotokous *A. m. capensis* workers within this locus. Another *A. m. scutellata* worker carried the *A. m. capensis Ethr* allele only. Sharing of alleles is not unexpected, as both subspecies form a stable hybrid zone, in which hybrid colonies can be observed (Hepburn and Crewe 1991). Such shared alleles also plausibly explain the successful crossing experiments that have been carried out between both subspecies in the past. Indeed, the segregation patterns observed in 28 previously published (back-) crosses between *A. m. capensis* and *A. m. scutellata* (Jordan et al. 2008; Oldroyd et al. 2014; Chapman et al. 2015) are in full agreement with the inferred balanced detrimental allele system at *Th* (supplementary table 3, Supplementary Material online).

Transcript Abundance Dynamics of *Th* Differ between Thelytokous and Arrhenotokous Workers

To test whether transcription patterns of *mycC*, *Th*, and *Ethr* show an association with the mode of parthenogenesis, we utilized previously established fat body transcriptome data (Aumer et al. 2018) and compared the gene expression between thelytokous *A. m. capensis* pseudoqueens (social parasitic lineage in NE South Africa) and arrhenotokous *A. m. scutellata* workers (NE South Africa), both sampled at four consecutive time points (3, 4, 7, and 8 days after emergence). In honey bees, the fat body is the key tissue involved in regulating ovary maturation (Amdam et al. 2011). After 8 days, all experimental *A. m. capensis* workers had fully activated ovaries, but the experimental *A. m. scutellata* workers were characterized by low levels of ovary development at all sampling time points (Aumer et al. 2018). Analyzing the gene expression over time, we found upregulation of *Th* by about three orders of magnitude after 8 days in the thelytokous *A. m. capensis* workers, while it remained at constantly low levels in the arrhenotokous *A. m. scutellata* workers (table 1). In addition, for both "early" (days 3/4) and "late" (days 7/8) time points, *Th* showed significantly higher expression in thelytokous (with both alleles Th_{Th} and Th_{ar} present) compared with arrhenotokous workers (table 1). Therefore, *Th* has highly thelytoky-specific transcript abundance dynamics. In addition to the reported coding sequence differences, the thelytoky-specific substitutions found in the 3′UTR of *Th* might have further regulatory influence (supplementary table 1, Supplementary Material online). The two other genes of interest in our comparison, *mycC* and *Ethr*, were not expressed at detectable levels in the fat bodies of the experimental workers at any sampled time point after emergence.

Discussion

In the past, both a single recessive locus (Lattorff et al. 2005; Jarosch et al. 2011) and multiple loci (Chapman et al. 2015; Wallberg et al. 2016) had been suggested to control thelytokous worker reproduction in *A. m. capensis*. However, the assumption of a recessive locus on chromosome 13 (Lattorff et al. 2005) was based on a false positive quantitative trait locus signal, most likely the result of the relatively small

Table 1. The Gene Expression of *Th* in Fat Bodies of Thelytokous and Arrhenotokous Workers during Sexual Maturation.

	Day 3/4	Day 7/8	Log 2-Fold Change (within groups)	Adjusted *P* Value (within groups)
Thelytokous workers (*A. m. capensis*)	**412.37 ± 162.86**	**808.81 ± 201.35**	**1.90**	**<0.01**
Arrhenotokous workers (*A. m. scutellata*)	**43.56 ± 6.50**	**31.41 ± 9.33**	**−0.46**	**0.92**
Log 2-fold change (between groups)	**3.24**	**4.68**		
Adjusted *P* value (between groups)	**<0.01**	**<0.01**		

NOTE.—The average expression value in the fat bodies (±SD; normalized counts) at the early (day 3/4) and the late (day 7/8) sampled time points after emergence of thelytokous *A. m. capensis* and arrhenotokous *A. m. scutellata* workers. Given are the log 2-fold change and adjusted *P* value for both pairwise comparisons within (day3/4 vs. day7/8) and between the phenotypic groups.

sample size and the low number of analyzed microsatellite markers (546 simple sequence repeat markers genome-wide). The subsequently identified transcription factor *gemini* within the candidate locus on chromosome 13 is a key regulating gene for worker reproduction in general (Jarosch et al. 2011; Jarosch-Perlow et al. 2018) but does not control the mode of parthenogenesis per se (Chapman et al. 2015; Aumer et al. 2017). The two studies suggesting a multilocus control suffered from methodological flaws such as insufficient experimental control and incorrect microsatellite data analyses (Chapman et al. 2015, reanalysis of their data supported single locus control [Aumer et al. 2017]), or lacking phenotyping of the mode of parthenogenesis and conflation of population structure (Wallberg et al. 2016). Hence, although the genetic basis of the thelytoky syndrome had been intensely studied, it remained to be controversially discussed and ambiguous.

The multifaceted basis of this study revealed that the genetic control of the thelytoky syndrome is regulated by a more complex genetic mechanism than previously assumed (Ruttner 1988; Lattorff et al. 2005; Jarosch et al. 2011; Chapman et al. 2015; Wallberg et al. 2016; Aumer et al. 2017). Indeed, thelytoky is controlled by a single locus (Ruttner 1988; Lattorff et al. 2005; Jarosch et al. 2011; Aumer et al. 2017), however the thelytoky allele (Th_{Th}) is not a recessive, but a dominant allele which operates in combination with a complementing arrhenotoky allele (Th_{ar}) that is under strong selection in *A. m. capensis*, to obtain workers that can reproduce thelytokously (Th_{Th}/Th_{ar}). It appears as though the arrhenotoky allele Th_{ar} found in *A. m. capensis* provides specific complementary functions similar to a "rescue allele" for Th_{Th}. In this case, Th_{Th} would need to interact with the specific Th_{ar} allele to result in thelytokous workers (Th_{Th}/Th_{ar}). Any other combination would distort the trait due to nonmatching alleles and would result in either nonfunctional (Th_{Th}/Th_{+}) or fertile arrhenotokous phenotypes (Th_{ar}/Th_{+}). This matches the mostly unsuccessful outcomes of attempted crossings of foreign honeybee subspecies with *A. m. capensis*, such as reported already in 1883 by Lord de Villiers, who never found fertile queens among the offspring of Italian queens that had been imported to South Africa (de Villiers 1883). Such a balanced detrimental allele system plausibly explains the stability of the hybrid zone between *A. m. capensis* and *A. m. scutellata* (Hepburn and Crewe 1991) and why thelytoky is not spreading into other honey bee populations, despite the various fitness benefits resulting from thelytokous worker reproduction.

In addition, due to the observed tight linkage, the single functional variant of *Ethr* that was detected in *A. m. capensis* may also be involved in the expression of the full thelytoky syndrome (Lattorff and Moritz 2013). *Ethr* encodes the receptor for the ecdysis-triggering hormone that regulates ecdysis and juvenile hormone synthesis in insects (Roller et al. 2010; Areiza et al. 2014; Meiselman et al. 2017) and is central for the regulation of larval and ovarian development. Even though no expression of *Ethr* was detected in the fat bodies of adult *A. m. capensis* pseudoqueens, the *A. m. capensis* functional *Ethr* variant may well be involved in the development of the queen-like traits in *A. m. capensis* workers during larval growth, including the high number of ovarioles (Ruttner 1977), the presence of a spermatheca (Ruttner 1988) and the production of queen-like pheromones (Hemmling et al. 1979; Crewe and Velthuis 1980; Okosun et al. 2017).

Similarly to *Ethr*, genotypes of the gene *mycC* in tight linkage with the Th_{Th} allele might be involved in the expression of the full thelytoky syndrome, as it is the orthologous gene of *ebony* of *Drosophila melanogaster*. *Ebony* is known to interact with dopamine signaling in insects (Borycz et al. 2002) and there are correlations between brain dopamine levels and ovary development (Harris and Woodring 1995), as well as between dopamine receptor expression in antennae and attraction to queen pheromones (Vergoz et al. 2009) in honey bee workers. Hence, in *A. m. capensis*, *mycC* might play a role in the behavioral changes from a social honeybee worker to a selfish social parasite (Moritz et al. 2002; Neumann and Hepburn 2002). Furthermore, *ebony* is known to be involved in pigmentation (Wittkopp and Beldade 2009) and thus, its ortholog *mycC* might drive the generally dark color of *A. m. capensis* workers (Ruttner 1977). On a broader level, the identified genetic architecture of thelytoky in honey bees may serve as a model for other eusocial species with thelytokous reproduction, in particular for novel ant model systems, such as *Platythyrea punctata* and the clonal raider ant *Ooceraea biroi*.

In conclusion, we show that thelytoky in *A. m. capensis* is controlled by a single heterozygous dominant locus located in a region under extreme selection. A single nonsynonymous SNP was inferred to be sufficient to lead to the striking phenotypic change from a social honey bee worker into a thelytokous social parasite. Thus, this study pertains to the rare examples of complex genetic structure influenced by strong purifying and balancing selection, and how a single

substitution affects major phenotypic changes (Wang et al. 2015; Ito et al. 2018), distorting sociality in the colony.

Materials and Methods

Am. capensis Mapping Population

Freshly emerged worker offspring of a single multiply mated *A. m. capensis* queen (colony at the Cape of Good Hope section of the Table Mountain National Park (34°14′45.0″S, 18°24′15.0″E)) provided a mapping population (see Aumer et al. 2017). A single individually labeled *A. m. capensis* worker (numbered tag on the thorax and paint marks on the abdomen) was introduced into a queenless mini-colony comprising about 1,000 freshly emerged queenless *A. m. scutellata* host workers, kept in Apidea mating nucs provided with sugar candy as well as honey-, pollen- and empty comb in bee-proof tents to prevent contamination or parasitism by other bees. These colonies ($n = 74$) were inspected at 2-day intervals to add more host workers, sugar candy or pollen as required. In such queenless host colonies, the *A. m. capensis* worker develops into a laying pseudoqueen within a few days (Johannsmeier 1983; Koeniger and Würkner 1992; Martin et al. 2002; Neumann and Hepburn 2002). Once larvae were emerging, but no later than 14 days after initiating the experiment, the *A. m. capensis* workers and all brood (eggs and/or larvae) were collected and stored at −20 °C until phenol–chloroform DNA extractions (Kirby 1957; Hunt and Page 1994). To confirm maternity and to determine the ploidy of the offspring (haploid = male, diploid = female), all *A. m. capensis* workers and their brood were genotyped at seven unlinked polymorphic microsatellite markers (A107, A079, A113, A014, A028, A088 and A035; Solignac et al. 2003), using a standard polymerase chain reaction (PCR) protocol (Solignac et al. 2003) and capillary sequencing (MegaBace 1000, GE Health Care, MegaBACE Fragment Profiler 1.3). Individuals with only one of the maternal alleles at all genotyped loci were classified as haploid male offspring. Individuals heterozygous with both maternal alleles at one or more loci were identified as diploid female offspring. Only individuals that produced reproducible genotypes for at least five of the seven marker loci (worker and all of its offspring) were included in the phenotype scoring, resulting in 21 thelytokous and 21 arrhenotokous workers that passed the genotyping threshold.

Whole Genome Sequencing and Variant Calling

The genomes of 71 *A. m. capensis* workers of the mapping population (42 with and 29 without confirmed phenotype) were sequenced on an Illumina HiSeq4000 Sequencing System (150 bp PE, TruSeq Nano DNA library preparation from 20 to 100 ng DNA, 350 bp insert size) to a coverage of ≥20× (raw reads are deposited in NCBI SRA, accession: PRJNA507348). We first performed quality assessment (fastqc v0.11.7 [bioinformatics.babraham.ac.uk/projects/fastqc]) of raw reads, followed by the removal of low quality reads and adapter contamination using skewer (Jiang et al. 2014) (v0.2.2, minimum window base quality of 20, minimum end-quality of 15, minimum length of 100, filter against degenerated N containing reads). The resultant reads were aligned to the *A. mellifera* reference genome (Amel_4.5 scaffold assembly) (Elsik et al. 2014) (GCA_000002195.1, scaffolds GL630009–GL635652, we used this assembly instead of the placed scaffolds to avoid known scaffold orientation problems) with BWA-MEM (Li and Durbin 2010) (v0.7.17-r1188). Alignments were processed and filtered (mapping quality ≥45, alignment length ≥80) in parallel using sambamba (Tarasov et al. 2015) (v0.6.7) and samtools (Li et al. 2009) (v1.6). Optical and PCR duplicates were removed with sambamba (Tarasov et al. 2015) markdup (v0.6.7) and the clumpify.sh script (BBMap [Bushnell 2014] package v37.86), respectively. Alignments were assessed using Qualimap2 (Okonechnikov et al. 2016) (v2.2.1) and any individual with a coverage of <20× was excluded from further analyses ($n = 1$).

Variants were identified using FreeBayes (Garrison and Marth 2012) (v1.1.0-54-g49413aa, –min-alternate-fraction 0.25 –min-alternate-total 2 –min-coverage 2). Variant calls from repetitive regions (tandem repeats finder [Benson 1999] v4.09) unreliable for read alignment were removed. The remaining variants were then decomposed and quality filtered (QUAL ≥ 20) and only biallelic SNPs that were present in at least 68 out of 70 individuals ($N = 7{,}238{,}467$) retained (vcflib [github.com/vcflib/vcflib, accessed January 23, 2018], vt [Tan et al. 2015] and vcftools [Danecek et al. 2011] [v0.1.15]). To determine the effect of nucleotide substitutions, we used SNPeff (Cingolani et al. 2012) (v4.3) to annotate all biallelic SNPs detected above. For this, we employed existing NCBI RefSeq annotations of the *A. mellifera* chromosome assembly (Amel_4.5, GCA_000002195.4) and lifted them over on our reference assembly (Amel_4.5, assembly GCA_000002195.1, scaffolds GL630009–GL635652, the most recent RefSeq annotations [GCA_000002195.4] are not available specifically for this scaffold assembly) using the software flo (github.com/wurmlab/flo, commit 8b7372d, accessed January 23, 2018).

Identification of the Thelytoky-Controlling Locus

We measured the genetic differentiation (F_{ST}) and absolute pairwise nucleotide differences (D_{XY}) between arrhenotokous and thelytokous workers to identify alleles cosegregating with the two phenotypes. To avoid pseudoreplication, a subset of 13 arrhenotokous and 13 thelytokous workers, representing 13 unique patrilines within each phenotypic group, was used to determine nucleotide diversity (π) as well as the observed heterozygosity for both thelytokous and arrhenotokous workers. We estimated these parameters across the genome in overlapping windows of 100 SNPs (50 SNP steps, maximal window size 20 kb) using VCFtools (Danecek et al. 2011) (v0.1.15) and published scripts (Martin et al. 2016) (github.com/simonhmartin/genomics_general). Using all individuals of the mapping population mirrored these results (data not shown).

The degree of linkage disequilibrium within the detected candidate region (scaffold Group1.23: 480,000–580,000 bp) was analyzed with HaploView (Barrett et al. 2005) (v4.2) using SNPs at 500-bp intervals. The minimum genotype threshold

was set to 75% and markers with rare subfamily-specific alleles were excluded from the analysis to avoid false negative signals as the squared Pearson coefficient of correlation (r^2) measurement was used.

Based on the SNP genotypes, for each coding sequence within the detected candidate region and the neighboring regions up- and downstream of the candidate region, the queen alleles were inferred for each gene following Mendelian inference (Estoup et al. 1995). The genotypes of all individuals of the mapping population ($n = 70$) were manually assessed to detect recombination events between the queen alleles.

To determine whether the detected region of divergence is part of an inversion, we analyzed the short read mapping data (above) using lumpy-SV (Layer et al. 2014) (v0.2.13), delly (Rausch et al. 2012) (v0.7.7), and samplot (https://github.com/ryanlayer/samplot). In addition, we sequenced genomic DNA of two thelytokous females from the parasitic lineage on a single Oxford Nanopore flowcell each (LSK108 1D ligation library preparation, R9.4.1 flowcell, raw reads are deposited in NCBI SRA, accession: PRJNA507349). Basecalling and adapter trimming was done with Albacore (OxfordNanoporeTechnologies) (v2.3.3) and porechop (github.com/rrwick/Porechop) (v0.2.3). One individual underperformed (~5× genomic coverage) and was thus excluded from further analyses. The other individual yielded 2,481,636 reads and 5,836,823,564 bases (~25× genomic coverage) of long reads in the "pass" quality category (mean read length = 2,352 bp, longest read = 104,006 bp, and read N50 = 4,532 bp). The trimmed reads were aligned to the Amel_4.5 scaffold assembly, and separately also to the recently available Amel_HAv3.1 assembly (GenBank GCA_03254395.1, Wallberg et al. 2018), with ngmlr (Sedlazeck et al. 2018) (v0.2.7) and structural variants were called with sniffles (Sedlazeck et al. 2018) (v1.0.9) with a threshold of minimum five reads supporting a variant.

Further, within the candidate region and its neighboring regions (400,000–460,000 and 600,000–660,000 bp) as well as within candidate genes, the number and proportion of heterozygous, homozygous, and fixed SNPs were evaluated. For candidate genes, the direction of selection was assessed using the McDonald–Kreitman test (McDonald and Kreitman 1991). Coding SNPs causing nonsynonymous substitutions as well as SNPs in the 5′UTRs and 3′UTRs of candidate genes were analyzed in more detail by assessing the individual genotypes to determine whether they are associated with the mode of parthenogenesis or fixed in the population. Based on these analyses, the haplotype sequences of the thelytoky and the arrhenotoky allele of *Th* as well as of *Ethr* were inferred.

In addition to our mapping population, we analyzed the genomes of four *A. m. capensis* individuals, offspring of social parasitic workers (thelytokous socially parasitic lineage of *A. m. capensis* in NE South Africa), collected from a parasitized *A. m. scutellata* colony in Pretoria (University of Pretoria, experimental farm) (sequencing methods and sequence analyses as outlined above, raw reads are deposited in NCBI SRA, accession: PRJNA507349). We also included published *A. mellifera* genome sequences into our analyses: 9 *A. m. carnica* (NCBI SRA accession: PRJNA216922; Harpur et al. 2014), 34 *A. m. scutellata* (NCBI SRA accessions: PRJNA294105, PRJNA237819; Harpur et al. 2014; Cridland et al. 2017; Wallberg et al. 2017), and 10 *A. m. yemenitica* workers (NCBI SRA accession: PRJNA294105; Harpur et al. 2014). For variant calling of the additional samples, the filtered biallelic variants detected in our mapping population were used as defined input, thus we only genotyped the additional individuals at polymorphic loci of the mapping population.

Protein Structure Analyses

The potential protein structure and function of the putative thelytoky gene (*Th*) were analyzed using the amino acid sequence (XP_397545.2) as input for Phyre2 (Kelley et al. 2015) (v2.0), PredictProtein (Yachdav et al. 2014), CCTOP (Dobson et al. 2015), and interproscan (Jones et al. 2014) (v68.0). Potential protein structures and functions were only reported if they were reliably reproducible with at least two programs. After the thelytoky and the arrhenotoky allele of *Th* were determined (based on variants that were present in all individuals of each phenotypic group), the corresponding amino acid sequences were translated from the distinct haplotype sequences and were analyzed with Phyre2 (Kelley et al. 2015) (v2.0) to model the tertiary structures and PredictProtein (Yachdav et al. 2014) to assess the number and the position of potential polynucleotide binding sites of each haplotype sequence separately, thus detecting potential changes between haplotypes. Similarly, the amino acid sequence of *Ethr* (XP_006570145.1) and the amino acid sequences derived from distinct haplotype sequences were analyzed.

Population Genetics

The previously identified patrilines of the mapping population, based on microsatellite data (Aumer et al. 2017), were reanalyzed based on 100 SNPs (spanning 2.995 bp: Group1.23: 506,639–509,446 bp) including and next to the thelytoky-associated variant (at 509,225 bp), following Mendelian inference (Estoup et al. 1995). The allele frequencies of the thelytoky (p_{Th}) and the arrhenotoky allele (p_{ar}) in the parental generation of the mapping population were determined. Based on these calculations, the relative fitness w and the selection coefficients s of the different genotypes (Th_{ar}/Th_{ar}, Th_{Th}/Th_{ar}, Th_{Th}/Th_{Th}) were assessed based on the following assumptions: heterozygous workers (Th_{Th}/Th_{ar}) reproduce thelytokously and can develop into pseudoqueens (Neumann and Moritz 2002), resulting in a maximum relative fitness ($w_{Th/ar} = 1$, $s_{Th/ar} = 0$). Since individuals that are homozygous for the thelytoky allele (Th_{Th}/Th_{Th}) were not detected, we could not exclude that Th_{Th}/Th_{Th} homozygous individuals are either lethal, infertile or experience significant fitness disadvantages. We therefore assumed a balanced detrimental allele system, where $w_{Th/Th} = 0$ and $s_{Th/Th} = 1$, resulting in the stable equilibrium frequency of $p_{ar} = s_{Th/Th}/(s_{Th/Th} + s_{ar/ar})$.

Allele Frequencies in Adult Drones

Because Th_{Th}/Th_{Th} homozygous individuals were not detected in our data set, we tested whether both Th_{ar} and

Th_{Th} are present in the drone offspring of the heterozygous mapping population's queen (Th_{Th}/Th_{ar}). Nineteen drones were collected in the colony and stored in ethanol at −20 °C. One hind-leg per drone was used for DNA extraction with 100 µl Chelex (Walsh et al. 1991) solution (5%) and 0.1 mg proteinase K, following the standard Chelex thermocycler protocol (Walsh et al. 1991). One microliter of total DNA per drone was subsequently used for PCR amplification of a 383 bp DNA fragment (508,685–509,068 bp) downstream of *Th* (one reaction contained 0.2 µM of each primer [primer 1: GGTTCTCTGTCGCTGCCTAT, primer 2: CCATGGTC AGTTCTTTGTTTCGT], 0.2 mM dNTPs, 0.25 units peqGOLD Taq-DNA-Polymerase and 10× reaction buffer S [both from Peqlab] in a total volume of 10 µl) following a standard PCR protocol (5 min at 94 °C, 38 cycles of 30 s at 94 °C, 30 s at 60 °C, 60 s at 72 °C, 15 min at 72 °C). PCR products were purified (QIAquick PCR Purification Kit, Qiagen) and Sanger sequenced (Eurofins Genomics, Germany) in one direction. Sequences were analyzed in BioEdit (Hall 1999) (v7.0.4) regarding a 3-bp deletion (508,880–50,882 bp) for which the presence (associated with Th_{Th}) or the absence (associated with Th_{ar}) was determined.

Thelytoky-Associated Transcript Abundance Dynamic Analysis

To assess transcript abundance dynamics of *Th*, *mycC*, and *Ethr* in thelytokous and arrhenotokous workers, we used previously published fat body RNAseq data (NCBI SRA accession: SRP135683; Aumer et al. 2018) from *A. m. capensis* workers of the socially parasitic lineage (thelytokous, obtained from infested *A. m. scutellata* colonies in NE South Africa) and *A. m. scutellata* workers (arrhenotokous, obtained from queenright *A. m. scutellata* colonies in NE South Africa). To follow the temporal transcriptomic changes during the first days after emergence, at least three replicate samples for both subspecies were taken at 3, 4, 7, and 8 days after starting the experimental cages (per time point six replicates for *A. m. capensis*, three replicates for *A. m. scutellata*). Usually, about 7–8 days after emergence, *A. m. capensis* workers have fully activated ovaries (Ruttner and Hesse 1981; Hepburn and Crewe 1991; Martin et al. 2002). For each time point and subspecies, three replicates were combined for full transcriptome sequencing (100 bp PE) on an Illumina HiSeq4000 Sequencing System (BGI, Hong Kong). Adapters were trimmed using Trimmomatic (Bolger et al. 2014) (v0.36) and all libraries (65,900,726 ± 6,190,602 [SD] reads per library [Aumer et al. 2018]) were aligned to the honeybee reference scaffold assembly Amel_4.5 (Elsik et al. 2014) (assembly GCA_000002195.1, scaffolds GL630009–GL635652) using HISAT2 (Kim et al. 2015) (v2.1.0) with default settings, followed by creation of a counts matrix including all samples and genes using HTSeq (Anders et al. 2015) (v0.7.1). To identify the differential expression levels of *Th*, *Ethr*, and *mycC* between thelytokous *A. m. capensis* and arrhenotokous *A. m. scutellata* workers, DeSeq2 (Love et al. 2014; Anders et al. 2015) (v1.20.0) was used on "early" replicated samples (days 3 and 4) and "late" replicated samples (days 7 and 8). Temporal transcript abundance dynamics of *Th* were determined using Next maSigPro (Nueda et al. 2014) (v1.52.0), which identifies genes that are significantly differentially expressed over time and between experimental groups. To assess whether both the thelytoky (Th_{Th}) and the arrhenotoky allele (Th_{ar}) of *Th* were present in the transcripts, we assessed genomic RNAseq alignments manually in IGV (Thorvaldsdóttir et al. 2013; Robinson et al. 2017). Here, the gene expression reanalysis was restricted to the candidate loci (*Th*, *Ethr*, and *mycC*); a complete transcriptome analysis can be found in Aumer et al. (2018).

Analyses of Previously Published Back-Crosses

To obtain a further additional and independent verification of our model of the genetic thelytoky control, we tested whether the balanced detrimental allele system is in agreement with the observed segregation patterns of previously published data sets on (back-) crosses between honey bee lineages of thelytokous and arrhenotokous laying workers (Lattorff et al. 2005; Jordan et al. 2008; Oldroyd et al. 2014; Chapman et al. 2015), following Mendelian inference (Estoup et al. 1995).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Joseph Thomas Colgan (University College Cork, Ireland), Yannick Wurm (Queen Mary University of London, United Kingdom), and Robert Paxton (Martin-Luther-University Halle-Wittenberg, Germany). This work was supported by grants from the Deutsche Forschungsgemeinschaft (to R.F.A.M.) and the National Research Foundation of South Africa (to C.W.W.P.).

Author Contributions

D.A., E.S., M.A., F.M., C.W.W.P., and R.F.A.M. designed the study and wrote the manuscript. D.A. and M.A. set up and sampled the *A. m. capensis* mapping population and the drones. D.A., E.S., and R.F.A.M. analyzed the genetic structure of the mapping population, the *A. m. capensis* social parasites from NE South Africa, and the other honey bee subspecies. E.S. performed and analyzed the Oxford Nanopore sequencing. D.A. and R.F.A.M. assessed the population genetics of *Th* in the natural *A. m. capensis* population. D.A., F.M., and C.W.W.P. obtained the RNAseq samples. D.A. analyzed the RNAseq data.

References

Allsopp MH. 1992. The capensis calamity. *S Afr Bee J.* 64:52–54.

Amdam GV, Fennern E, Havukainen H. 2011. Vitellogenin in honey bee behavior and lifespan. In: Galicia CG, Eisenhardt D, Giurfa M, editors. Honeybee neurobiology and behavior. Dordrecht (The Netherlands): Springer. p. 17–29.

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2): 166–169.

Areiza M, Nouzova M, Rivera-Perez C, Noriega FG. 2014. Ecdysis triggering hormone ensures proper timing of juvenile hormone biosynthesis in pharate adult mosquitoes. *Insect Biochem Mol Biol*. 54:98–105.
Aumer D, Allsopp MH, Lattorff HMG, Moritz RFA, Jarosch-Perlow A. 2017. Thelytoky in Cape honeybees (*Apis mellifera capensis*) is controlled by a single recessive locus. *Apidologie* 48(3): 401–410.
Aumer D, Mumoki FN, Pirk CWW, Moritz RFA. 2018. The transcriptomic changes associated with the development of social parasitism in the honeybee *Apis mellifera capensis*. *Naturwissenschaften* 105(3–4): 22.
Barrett JC, Fry B, Maller J, Daly MJ. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2): 263–265.
Baudry E, Kryger P, Allsopp M, Koeniger N, Vautrin D, Mougel F, Cornuet J-M, Solignac M. 2004. Whole-genome scan in thelytokous-laying workers of the Cape honeybee (*Apis mellifera capensis*): central fusion, reduced recombination rates and centromere mapping using half-tetrad analysis. *Genetics* 167(1): 243–252.
Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27(2): 573–580.
Blobel G, Dobberstein B. 1975. Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J Cell Biol*. 67(3): 835–851.
Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114–2120.
Borycz J, Borycz JA, Loubani M, Meinertzhagen IA. 2002. *tan* and *ebony* genes regulate a novel pathway for transmitter metabolism at fly photoreceptor terminals. *J Neurosci*. 22(24): 10549–10557.
Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Report No.: LBNL-7065E. Available from: https://sourceforge.net/projects/bbmap/
Chapman NC, Beekman M, Allsopp MH, Rinderer TE, Lim J, Oxley PR, Oldroyd BP. 2015. Inheritance of thelytoky in the honey bee *Apis mellifera capensis*. *Heredity* 114(6): 584–592.
Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6(2): 80–92.
Crewe RM, Velthuis HHW. 1980. False queens: a consequence of mandibular gland signals in worker honeybees. *Naturwissenschaften* 67(9): 467–469.
Cridland JM, Tsutsui ND, Ramírez SR. 2017. The complex demographic history and evolutionary origin of the western honey bee, *Apis mellifera*. *Genome Biol Evol*. 9(2): 457–472.
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15): 2156–2158.
de Villiers JH. 1883. Cape bees and "animal intelligence." *Nature* 28:5–6.
Dobson L, Reményi I, Tusnády GE. 2015. CCTOP: a Consensus Constrained TOPology prediction web server. *Nucleic Acids Res*. 43(W1): W408–W412.
Elsik CG, Worley KC, Bennett AK, Beye M, Camara F, Childers CP, de Graaf DC, Debyser G, Deng J, Devreese B, et al. 2014. Finding the missing honey bee genes: lessons learned from a genome upgrade. *BMC Genomics* 15:86.
Estoup A, Scholl A, Pouvreau A, Solignac M. 1995. Monoandry and polyandry in bumble bees (Hymenoptera; Bombinae) as evidenced by highly variable microsatellites. *Mol Ecol*. 4(1): 89–93.
Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv [q-bio.GN] [Internet]. Available from: http://arxiv.org/abs/1207.3907
Greeff JM. 1996. Effects of thelytokous worker reproduction on kin-selection and conflict in the Cape honeybee, *Apis mellifera capensis*. *Philos Trans R Soc Lond B Biol Sci*. 351(1340): 617–625.
Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 41:95–98.
Hamilton WD. 1964. The genetical evolution of social behaviour. I. *J Theor Biol*. 7(1): 1–16.
Harpur BA, Kent CF, Molodtsova D, Lebon JMD, Alqarni AS, Owayss AA, Zayed A. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc Natl Acad Sci U S A*. 111(7): 2614–2619.
Harris JW, Woodring J. 1995. Elevated brain dopamine levels associated with ovary development in queenless worker honey bees (*Apis mellifera* L.). *Comp Biochem Physiol C Pharmacol Toxicol Endocrinol*. 111(2): 271–279.
Hemmling C, Koeniger N, Ruttner F. 1979. Quantitative Bestimmung der 9-Oxodecansäure im Lebenszyklus der Kapbiene (*Apis mellifera capensis* ESCHOLTZ). *Apidologie* 10(3): 227–240.
Hepburn HR, Crewe RM. 1991. Portrait of the Cape honeybee, *Apis mellifera capensis*. *Apidologie* 22(6): 567–580.
Hepburn HR, Magnuson P, Herbert L, Whiffler LA. 1991. The development of laying workers in field colonies of the Cape honey bee. *J Apic Res*. 30(2): 107–112.
Holmes MJ, Oldroyd BP, Allsopp MH, Lim J, Wossler TC, Beekman M. 2010. Maternity of emergency queens in the Cape honey bee, *Apis mellifera capensis*. *Mol Ecol*. 19(13): 2792–2799.
Hunt GJ, Page RE Jr. 1994. Linkage analysis of sex determination in the honey bee (*Apis mellifera*). *Mol Gen Genet*. 244(5): 512–518.
Ito K, Kidokoro K, Katsuma S, Sezutsu H, Uchino K, Kobayashi I, Tamura T, Yamamoto K, Mita K, Shimada T, et al. 2018. A single amino acid substitution in the *Bombyx*-specific mucin-like membrane protein causes resistance to *Bombyx mori* densovirus. *Sci Rep*. 8(1): 7430.
Jackson BC, Campos JL, Zeng K. 2015. The effects of purifying selection on patterns of genetic differentiation between *Drosophila melanogaster* populations. *Heredity* 114(2): 163–174.
Jarosch A, Stolle E, Crewe RM, Moritz RFA. 2011. Alternative splicing of a single transcription factor drives selfish reproductive behavior in honeybee workers (*Apis mellifera*). *Proc Natl Acad Sci U S A*. 108(37): 15282–15287.
Jarosch-Perlow A, Yusuf AA, Pirk CWW, Crewe RM, Moritz RFA. 2018. Control of mandibular gland pheromone synthesis by alternative splicing of the CP-2 transcription factor gemini in honeybees (*Apis mellifera carnica*). *Apidologie* 49(4): 450–458.
Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:182.
Johannsmeier MF. 1983. Experiences with the Cape bee in the Transvaal. *S Afr Bee J*. 55:130–138.
Jones P, Binns D, Chang H-Y, Fraser M, Li W, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240.
Jordan LA, Allsopp MH, Beekman M, Wossler TC, Oldroyd BP. 2008. Inheritance of traits associated with reproductive potential in *Apis mellifera capensis* and *Apis mellifera scutellata* workers. *J Hered*. 99(4): 376–381.
Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 10(6): 845–858.
Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 12(4): 357–360.
Kirby KS. 1957. A new method for the isolation of deoxyribonucleic acids; evidence on the nature of bonds between deoxyribonucleic acid and protein. *Biochem J*. 66(3): 495–504.
Kirkpatrick M. 2010. How and why chromosome inversions evolve. *PLoS Biol*. 8:e10000501.
Koeniger N, Würkner W. 1992. Die Kap Honigbienen (*Apis mellifera capensis*) Natürliche Verbreitung und die Schwierigkeiten der Haltung unter unseren klimatischen Bedingungen. *Die Biene* 138:583–588.
Lattorff HMG, Moritz RFA. 2013. Genetic underpinnings of division of labor in the honeybee (*Apis mellifera*). *Trends Genet*. 29(11): 641–648.
Lattorff HMG, Moritz RFA, Crewe RM, Solignac M. 2007. Control of reproductive dominance by the thelytoky gene in honeybees. *Biol Lett*. 3(3): 292–295.
Lattorff HMG, Moritz RFA, Fuchs S. 2005. A single locus determines thelytokous parthenogenesis of laying honeybee workers (*Apis mellifera capensis*). *Heredity* 94(5): 533–537.

Layer RM, Chiang C, Quinlan AR, Hall IM. 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 15(6): R84.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26(5): 589–595.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078–2079.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15(12): 550.
Martin SH, Möst M, Palmer WJ, Salazar C, McMillan WO, Jiggins FM, Jiggins CD. 2016. Natural selection and genetic diversity in the butterfly *Heliconius melpomene*. *Genetics* 203(1): 525–541.
Martin SJ, Beekman M, Wossler TC, Ratnieks FLW. 2002. Parasitic Cape honeybee workers, *Apis mellifera capensis*, evade policing. *Nature* 415(6868): 163–165.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila melanogaster*. *Nature* 351(6328): 652–654.
Meiselman M, Lee SS, Tran R-T, Dai H, Ding Y, Rivera-Perez C, Wijesekera TP, Dauwalder B, Noriega FG, Adams ME. 2017. Endocrine network essential for reproductive success in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 114(19): E3849–E3858.
Moritz RFA. 1986. Two parthenogenetical strategies of laying workers in populations of the honeybee, *Apis mellifera* (Hymenoptera: Apidae). *Entomol Gen*. 11(3–4): 159–164.
Moritz RFA, Crewe RM, Hepburn HR. 2002. Queen avoidance and mandibular gland secretion of honeybee workers (*Apis mellifera* L.). *Insectes Soc*. 49(1): 86–91.
Neumann P, Hepburn R. 2002. Behavioural basis for social parasitism of Cape honeybees (*Apis mellifera capensis*). *Apidologie* 33(2): 165–192.
Neumann P, Moritz R. 2002. The Cape honeybee phenomenon: the sympatric evolution of a social parasite in real time? *Behav Ecol Sociobiol*. 52:271–281.
Nueda MJ, Tarazona S, Conesa A. 2014. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics* 30(18): 2598–2602.
Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32(2): 292–294.
Okosun OO, Pirk CWW, Crewe RM, Yusuf AA. 2017. Glandular sources of pheromones used to control host workers (*Apis mellifera scutellata*) by socially parasitic workers of *Apis mellifera capensis*. *J Insect Physiol*. 102:42–49.
Oldroyd BP, Allsopp MH, Roth KM, Remnant EJ, Drewell RA, Beekman M. 2014. A parent-of-origin effect on honeybee worker ovary size. *Proc Biol Sci*. 281(1775): 20132388.
Onions GW. 1914. South African "fertile" worker bees. *Agric J Union S Afr*. 7:44–46.
Pirk CWW, Human H, Crewe RM, vanEngelsdorp D. 2014. A survey of managed honey bee colony losses in the Republic of South Africa–2009 to 2011. *J Apic Res*. 53(1): 35–42.
Pirk CWW, Neumann P, Hepburn RH. 2002. Egg laying and egg removal by workers are positively correlated in queenright Cape honeybee colonies (*Apis mellifera capensis*). *Apidologie* 33(2): 203–211.
Plettner E, Slessor KN, Winston ML, Robinson GE, Page RE. 1993. Mandibular gland components and ovarian development as measures of caste differentiation in the honey bee (*Apis mellifera* L.). *J Insect Physiol*. 39(3): 235–240.
Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28(18): i333–i339.
Robinson JT, Thorvaldsdóttir H, Wenger AM, Zehir A, Mesirov JP. 2017. Variant review with the integrative genomics viewer. *Cancer Res*. 77(21): e31–e34.
Roller L, Zitnanová I, Dai L, Simo L, Park Y, Satake H, Tanaka Y, Adams ME, Zitnan D. 2010. Ecdysis triggering hormone signaling in arthropods. *Peptides* 31(3): 429–441.
Ruttner F. 1977. The problem of the Cape bee (*Apis mellifera capensis* ESCHOLTZ): parthenogenesis—size of population—evolution. *Apidologie* 8(3): 281–294.
Ruttner F. 1988. Biogeography and taxonomy of honeybees. Heidelberg (Germany): Springer Verlag.
Ruttner F, Hesse B. 1981. Rassenspezifische Unterschiede in Ovarentwicklung und Eiablage von weisellosen Arbeiterinnen der Honigbiene *Apis mellifera* L. *Apidologie* 12(2): 159–183.
Sakagami SF. 1958. The false-queen: fourth adjustive response in dequeened honeybee colonies. *Behaviour* 13(3–4): 280–295.
Schaid DJ, Chen W, Larson NB. 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet*. 19(8): 491–504.
Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*. 15(6): 461–468.
Slessor KN, Kaminski L-A, King GGS, Borden JH, Winston ML. 1988. Semiochemical basis of the retinue response to queen honey bees. *Nature* 332(6162): 354–356.
Solignac M, Vautrin D, Loiseau A, Mougel F, Baudry E, Estoup A, Garnery L, Haberl M, Cornuet J-M. 2003. Five hundred and fifty microsatellite markers for the study of the honeybee (*Apis mellifera* L.) genome. *Mol Ecol Notes*. 3(2): 307–311.
Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philos Trans R Soc Lond B Biol Sci*. 372(1736): 20160455.
Tan A, Abecasis GR, Kang HM. 2015. Unified representation of genetic variants. *Bioinformatics* 31(13): 2202–2204.
Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. 2015. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* 31(12): 2032–2034.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 14(2): 178–192.
Tribe GD. 1983. What is the Cape bee? *S Afr Bee J*. 55:77–87.
Vergoz V, McQuillan HJ, Geddes LH, Pullar K, Nicholson BJ, Paulin MG, Mercer AR. 2009. Peripheral modulation of worker bee responses to queen mandibular pheromone. *Proc Natl Acad Sci U S A*. 106(49): 20930–20935.
Verma S, Ruttner F. 1983. Cytological analysis of the thelytokous parthenogenesis in the Cape honeybee (*Apis mellifera capensis* ESCHOLTZ). *Apidologie* 14(1): 41–57.
Wallberg A, Bunikis I, Pettersson OV, Mosbech M-B, Childers AK, Evans JD, Mikheyev AS, Robertson HM, Robinson GE, Webster MT. 2018. A hybrid *de novo* genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *bioRxiv*. doi:10.1101/361469.
Wallberg A, Pirk CW, Allsopp MH, Webster MT. 2016. Identification of multiple loci associated with social parasitism in honeybees. *PLoS Genet*. 12(6): e1006097.
Wallberg A, Schöning C, Webster MT, Hasselmann M. 2017. Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLoS Genet*. 13(5): e1006792.
Walsh PS, Metzger DA, Higuchi R. 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10(4): 506–513.
Wang H, Studer AJ, Zhao Q, Meeley R, Doebley JF. 2015. Evidence that the origin of naked kernels during maize domestication was caused by a single amino acid substitution in tga1. *Genetics* 200(3): 965–974.
Wittkopp PJ, Beldade P. 2009. Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. *Semin Cell Dev Biol*. 20(1): 65–71.
Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, Hönigschmid P, Schafferhans A, Roos M, Bernhofer M, et al. 2014. PredictProtein—an open resource for online prediction of protein structural and functional features. *Nucleic Acids Res*. 42(W1): W337–W343.