Insect Molecular Biology

Insect Molecular Biology (2016) **25**(3), 191–201

doi: 10.1111/imb.12210

The *transformer* genes in the fig wasp *Ceratosolen solmsi* provide new evidence for duplications independent of complementary sex determination

L.-Y. Jia*, J.-H. Xiao*, T.-L. Xiong*, L.-M. Niu† and D.-W. Huang*‡

**Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China;* †*Environment and Plant Protection Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan, China; and* ‡*Plant Protection College, Shandong Agricultural University, Tai'an, China*

Abstract

***Transformer* (*tra*) is the key gene that turns on the sex-determination cascade in *Drosophila melanogaster* and in some other insects. The honeybee *Apis mellifera* has two duplicates of *tra*, one of which (complementary sex determiner, *csd*) is the primary signal for complementary sex-determination (CSD), regulating the other duplicate (feminizer). Two *tra* duplicates have been found in some other hymenopteran species, resulting in the assumption that a single ancestral duplication of *tra* took place in the Hymenoptera. Here, we searched for *tra* homologues and pseudogenes in the Hymenoptera, focusing on five newly published hymenopteran genomes. We found three *tra* copies in the fig wasp *Ceratosolen solmsi*. Further evolutionary and expression analyses also showed that the two duplicates (*Csoltra-B* and *Csoltra-C*) are under positive selection, and have female-specific expression, suggesting possible sex-related functions. Moreover, Aculeata species exhibit many pseudogenes generated by lineage-specific duplications. We conclude that phylogenetic reconstruction and pseudogene screening provide novel evidence supporting the hypothesis of independent duplications rather an ancestral origin of multiple *tra* paralogues in the Hymenoptera. The case of *C. solmsi* is the first example of a non-CSD species with duplicated *tra*, contrary to the previous assumption that derived *tra* paralogues function as the CSD locus.**

Keywords: complementary sex-determination, inbreeding, pseudogene, alternative splicing, gene duplication.

First published online 8 January 2016.

Correspondence: Jin-Hua Xiao or Da-Wei Huang, Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; e-mails: xiaojh@ioz.ac.cn; huangdw@ioz.ac.cn

Introduction

The protein Transformer (Tra) is a serine/arginine (SR)-related protein, which possesses SR domains, mediates protein–protein interactions and is involved in pre-mRNA splicing (Lynch & Maniatis, 1996). In *Drosophila melanogaster*, the *transformer* gene *Dmtra* occupies a critical position in the sex-determination cascade, with the protein Sex-lethal (SXL) regulating its functional female-specific splicing. Female-specific Tra acts in conjunction with some other proteins to regulate the RNA splicing of *doublesex* and other genes to ultimately control somatic sexual differentiation (McKeown *et al.*, 1988; Belote *et al.*, 1989; Hedley & Maniatis, 1991). Since the identification of *D. melanogaster tra* orthologues in many other insect species, *tra* has received much attention (Verhulst *et al.*, 2010). The *tra* gene has been found to be widely distributed amongst insects, and its function as a sex-determining gene may be conserved (Geuverink & Beukeboom, 2014).

The *tra* gene has only been found to be duplicated in the Hymenoptera, with the duplicated paralogue [the *complementary sex determiner* (*csd*) gene] being crucial in complementary sex-determination (CSD) in *Apis mellifera*. As the best-studied genetic example of the CSD system, *Apis mellifera* has two homologues of *tra*, both of which are recruited into the sex-determination cascade: one termed *feminizer* (*fem*), and the other being the *csd* gene (Beye *et al.*, 2003; Hasselmann *et al.*, 2008). Csd is also an SR-related protein that has marked variations in the proline-rich domain, and initiates female-specific

© 2016 The Royal Entomological Society

splicing of *fem* by its allelic composition (Beye *et al.*, 2003). The principle of CSD is that individuals with heterozygous *csd* alleles become diploid females. Males are thus normally haploid, but in inbred populations diploid males that have homozygous *csd* genotypes can occur (Whiting, 1939, 1943). To date, the occurrence of *tra* in the Aculeata hymenopteran clade (such as bees and ants) has been identified, with duplicates of *tra* orthologues known to be present in several species (Schmieder *et al.*, 2012; Privman *et al.*, 2013; Geuverink & Beukeboom, 2014). We here use the terminology of Privman *et al.* (2013) and refer to these duplicates as *tra-A* and *tra-B*. However, it is unclear if *tra* paralogues in these hymenopterans are generated by one or a few ancestral duplications or by multiple independent duplications. Furthermore, previously reported data suggested that only CSD species have duplicated *tra* genes (Schmieder *et al.*, 2012). Besides, studies of *tra* in the Chalcidoidea and Symphyta are few (Schmieder *et al.*, 2012; Privman *et al.*, 2013; Koch *et al.*, 2014).

Although CSD is a valid explanation for sex determination in many species, and thought to be ancient in the Hymenoptera (Asplen *et al.*, 2009; Schmieder *et al.*, 2012), the failure to produce diploid males in inbred populations in some species, especially in parasitoid wasps, suggests that other mechanisms are present (Beukeboom, 1995). One of the best-studied examples is the parasitoid *Nasonia vitripennis*, in which the *csd* and any other *tra* paralogue genes are absent (Werren *et al.*, 2010). In *N. vitripennis*, the sex determination is controlled by the maternal input of *tra* mRNA and the specific zygotic *tra* transcription, known as maternal effect genomic imprinting sex-determination (Beukeboom *et al.*, 2007; van de Zande & Verhulst, 2014). However, in other parasitoids of the Cynipoidea and Chalcidoidea, which lack CSD, the mechanism of sex determination remains to be clarified (Beukeboom *et al.*, 2007; Asplen *et al.*, 2009). The evolutionary trajectory of *tra* within the parasitoid Hymenoptera is also unclear.

Recently sequenced hymenopteran genomes enable comparative investigation of *tra* in the Hymenoptera. Here, we searched the genomes of several hymenopteran species for *tra* homologues and pseudogenes. The investigated species are three chalcids (Apocrita; Parasitica: *Ceratosolen solmsi*, *Copidosoma floridanum* and *Trichogramma pretiosum*), a single bee (Halictidae; Halictini: *Lasioglossum albipes*) and a Symphyta outgroup species (*Orussus abietinus*). We not only elucidated the evolutionary trajectory of the *tra* gene and its pseudogenes in hymenopteran insects, but also focused on the unexpected presence of three *tra* gene duplicates in the fig wasp *Ce. solmsi*, the first report of *tra* duplications in a non-CSD hymenopteran. We present findings on the evolution of the two exclusive *tra* genes, including information on their origin and evidence showing that these genes are under positive selection. We speculate that the proteins encoded by the duplicated *tra* genes in *Ce. solmsi* with newly evolved C-terminal domains may have elicited some new sex-related functions.

Results

Identification and phylogenetic analyses of hymenopteran tra *homologues*

We identified the potentially functional orthologues of the *tra* gene for five newly available hymenopteran genomes from one bee, three chalcid and one Symphyta species (Table S1). Two *tra* homologues are present in the bee *L. albipes*, which undergoes CSD (Smith, 1983; Kukuk & May, 1990), whereas the Symphyta outgroup species *O. abietinus* has only one *tra* gene. Of the three non-CSD chalcid species (Grbic *et al.*, 1992; Stouthamer & Kazmer, 1994; Peng *et al.*, 2014), *Ce. solmsi* has three *tra* homologues (*Csoltra-A*, *Csoltra-B* and *Csoltra-C* with GenBank accession numbers KP736170, KP757104 and KP757105), whereas the other two (*T. pretiosum* and *Co. floridanum*) possess only a single copy. It is noteworthy that owing to incomplete genome assemblies, the predicted *Coflotra* gene (*tra* of *Co. floridanum*) is localized on two different scaffolds. Furthermore, the sequence of *Tpretra* (*tra* of *T. pretiosum*; only exons 1–3 were identified) is partial; in *L. albipes*, with the exception of the two *tra* homologues, there are *tra* copies with partial coding sequences (data not shown). All of the partial sequences of *T. pretiosum* and *L. albipes* were excluded from further analyses. The organization of *tra* in these species therefore requires further investigation.

The domains reported in Tra/Fem are well known in the Diptera and in the Hymenoptera (Hediger *et al.*, 2010; Saccone *et al.* 2011; Schmieder *et al.*, 2012). All of our newly annotated Tra/Fem proteins (except for CsolTra-C, the C-terminal of which varies considerably) share four main domains as found previously (Verhulst *et al.*, 2010; Geuverink & Beukeboom, 2014): a sex-determiner protein amino-terminal (SDP_N) domain, a autoregulation domain (CAM) domain, a SR domain and a Pro-rich domain. In *Ce. solmsi*, the homologous amino acid sequences of *Csoltra-B* and *Csoltra-C*, including the SDP_N domain and the CAM domain (auto-regulation domain), are similar (65–67.2% sequence similarity) to CsolTra-A proteins. We found no significant similarities to other known proteins.

Protein sequences of 28 hymenopteran *tra* genes were aligned by BALI-PHY (Suchard & Redelings, 2006) and used for phylogenetic analyses (the sequences used are listed in Tables S1 and S2). Despite the different methods of phylogenetic analyses, trees with similar topologies were produced (trees reconstructed by BALI-PHY

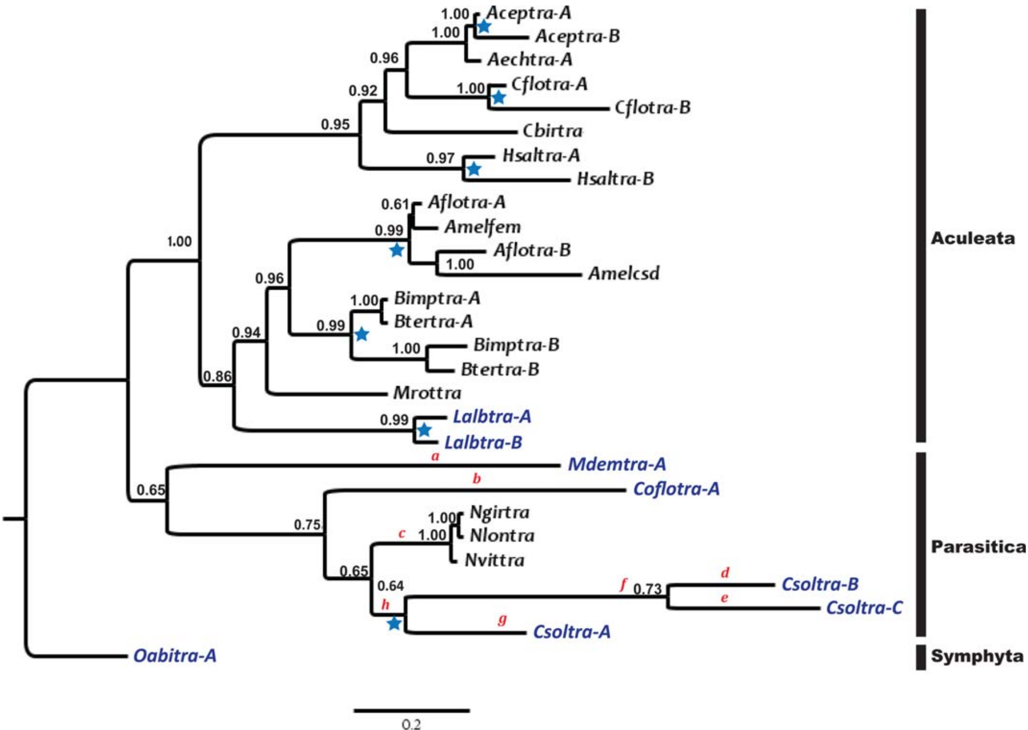

Figure 1. Bayesian phylogenetic analysis based on the deduced amino acid sequences of *transformer* (*tra*) homologous genes in the Hymenoptera using BAli-Phy. Posterior probability support is indicated for each node. *tra* homologues annotated in this study are coloured in blue. The branches *a* to *h* are the lineages tested for evolutionary selection with the branch and branch-site models using paml software (Table 1); a blue star beside a clade node indicates a predicted independent duplication event. The blue star in the *Bombus tra* clade highlights the duplication that generated the ancestral pseudogene of *Bombus tra-A* before the divergence of the two *Bombus* species (ancestral pseudogene of *Bimp_traφ1* and *Bter_traφ1*, see Fig. 2 and Tables S3, S4). The scale bar represents the estimated number of substitutions per site. Abbreviations: *fem*, *feminizer*; *Acep*, *Atta cephalotes*; *Aech*, *Acromyrmex echinatior*; *Cflo*, *Camponotus floridanus*; *Cbir*, *Cerapachys biroi*; *Hsal*, *Harpegnathos saltator*; *Aflo*, *Apis florea*; *Amel*, *Apis mellifera*; *Bimp*, *Bombus impatiens*; *Bter*, *Bombus terrestris*; *Mrot*, *Megachile rotundata*; *Lalb*, *Lasioglossum albipes*; *Mdem*, *Microplitis demolitor*; *Coflo*, *Copidosoma floridanum*; *Ngir*, *Nasonia giraulti*; *Nlon*, *Nasonia longicornis*; *Nvit*, *Nasonia vitripennis*; *Csol*, *Ceratosolen solmsi*; *Oabi*, *Orussus abietinus*.

are shown in Fig. 1, whereas tree topologies inferred from Bayesian inference using MrBayes (Huelsenbeck & Ronquist, 2001) and that inferred from maximum likelihood using PhyML (Guindon *et al*., 2010) are presented in Fig. S1). Our results are consistent with previously published *tra* phylogenies, and suggest that *tra* has been duplicated repeatedly and independently in different species (such as in *Ce. solmsi* and ants) or genera (such as in honeybees and bumblebees) (Koch *et al*., 2014). The *tra* in *Microplitis demolitor* shows a closer relationship to Chalcidoidea *tra*, which is consistent with the closer relationship of the Ichneumonoidea with the Chalcidoidea than with the Aculeata (Mao *et al*., 2015). The single copy of the *tra* gene in the basal lineage, *O. abietinus*, suggests that the independent duplication of *tra* may have occurred after the branching off of the Symphyta. More studies on Symphyta species are needed to test this hypothesis. The paralogous *tra* gene appears to be absent in most parasitoids. By contrast, three *tra* homologues are present in *Ce. solmsi*, with full coding sequences and no premature stop codons. Our phylogenetic analysis shows that the three *tra* genes in *Ce. solmsi* form a single lineage, with *Csoltra-B* and *Csoltra-C* being most closely related. In addition, the 5′-untranslated regions (5′-UTRs) of *Csoltra-B* and *Csoltra-C* are considerably similar to each other (86.81%). These results suggest two possible *tra* duplication events in *Ce. solmsi*, with the gene duplication that contributed to the generation of *Csoltra-B* and *Csoltra-C* being the most recent event.

Pseudogenes in Aculeata indicate repeated duplications of tra *in the Hymenoptera*

Koch *et al*. (2014) searched for *tra* pseudogenes in one honeybee, two bumblebee and two ant species. We expanded this to include 19 hymenopteran genomes, covering representatives of the Formicidae, Apoidea, Chalcidoidea, Braconidae and Symphyta. We obtained all of the putative pseudogene fragments that map to each exon of the *tra* genes (Table S3). Our study provides the most thorough search of pseudogenes in the Hymenoptera to date (Schmieder *et al*., 2012; Koch *et al*., 2014), and enables us to conclude that pseudogenes are only present in bees and ants, and are absent from the other lineages examined (Fig. 2).

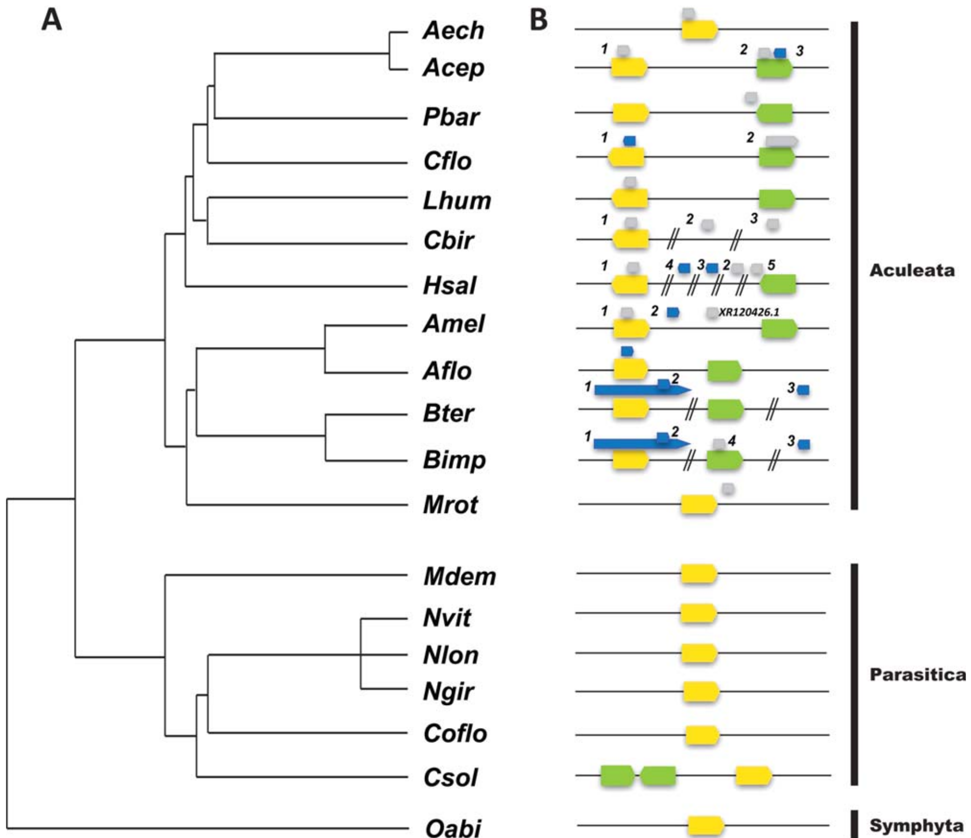

Figure 2. Diagram showing the relative genomic positions and orientations of the homologues and pseudogene fragments of *feminizer* (*fem*)/ *complementary sex determiner* in selected Hymenoptera. (A) The phylogeny of the hymenopteran species included in the present study as inferred from earlier studies (Danforth *et al.*, 2013; Heraty *et al.*, 2013; Moreau & Bell, 2013; Mao *et al.*, 2015). (B) Relative genome position and orientations of inferred genes and pseudogenes are represented by coloured boxes with arrow heads. *Transformer-A* (*tra-A*)/*Apis mellifera fem* are shown by yellow arrow heads, whereas *tra-B*/C are shown in green. Small grey arrow heads above indicate the pseudogenes with an early predicted origin; blue arrow heads indicate pseudogenes with a more recent origin (details in Table S4). All pseudogenes are numbered as presented in Table S3. Double slashes separate distinct scaffolds. Abbreviations: *Acep*, *Atta cephalotes*; *Aech*, *Acromyrmex echinatior*; *Cflo*, *Camponotus floridanus*; *Cbir*, *Cerapachys biroi*; *Hsal*, *Harpegnathos saltator*; *Aflo*, *Apis florea*; *Amel*, *Apis mellifera*; *Bimp*, *Bombus impatiens*; *Bter*, *Bombus terrestris*; *Mrot*, *Megachile rotundata*; *Mdem*, *Microplitis demolitor*; *Coflo*, *Copidosoma floridanum*; *Ngir*, *Nasonia giraulti*; *Nlon*, *Nasonia longicornis*; *Nvit*, *Nasonia vitripennis*; *Csol*, *Ceratosolen solmsi*; *Oabi*, *Orussus abietinus*; *Pbar*, *Pogonomyrmex barbatus*; *Lhum*, *Linepithema humile*.

It is noteworthy that almost all of the ants and bees that we examined possessed at least one pseudogene fragment homologous to one to three exons of *tra-A/fem* (Table S3). We therefore traced the evolution of the internal duplication of these exons throughout the phylogeny of bees and ants, by estimating the synonymous per site substitutions (d_S) of the pseudogene fragments and of the *tra* genes within each species. We compared the d_S between the pseudogene and its parent gene, and the d_S between the most related *tra* gene pairs. The most related *tra* genes were either orthologous *tra* genes in the most closely related species, or paralogous *tra* pairs arising from independent duplication in each species. These pairs were chosen according to the results of our phylogenetic analyses (Fig. 1), such as *fem*/*tra-A* and *csd*/*tra-B* in *Apis* and *Bombus*, paralogues in each Formicidae species and *L. albipes*, and *tra-A* genes in species with only one copy of *tra*. To our surprise, of the examined 20 pseudogenes which are homologous to one to three exons of *tra*-A*/fem*, 13 showed a lower level of d_S divergence between pseudogenes and their parent genes compared with the d_S between the most related *tra* homologous genes (Table S4). In other words, most of the pseudogene fragments homologous to *tra* exons 1–3 had duplicated recently (Fig. 2), with these duplication events having occurred repeatedly and independently.

We also found analogous pseudogene regions with almost complete gene structure in several species. Such pseudogenes are present in bumblebee genomes (*Bombus terrestris* and *Bombus impatiens*; Schmieder *et al.*, 2012; Koch *et al.*, 2014). For this study, we explored both of these genomes and detected synteny between pseudogenes *Bter_tra*φ*1* and *Bimp_tra*φ*1*, providing further evidence that these pseudogenes were present in the common ancestor of both bumblebee

Table 1. Tests for positive selection on *transformer* (*tra*) homologues using the PAML branch model and branch-site model A

Model parameters	*CsoltraA*	*CsoltraB*	*CsoltraC*	*CsoltraB/C*	*Csoltra* genes	*Nasonia tra* genes	*Coflotra*	*Mdemtra*
Branch models								
B0: $\omega 0$ ($\omega 1 = 1$)	0.18	0.18	0.18	0.18	0.18	0.18	0.18	0.19
BA: $\omega 0$, $\omega 1$	0.19, 0.07	0.18, 0.28	0.18, 0.30	0.18, 0.41	0.19, 0.02	0.20, 0.04	0.19, 0.08	0.19, 0.02
Branch models on N-terminal*								
B'0: $\omega 0$ ($\omega 1 = 1$)	0.15	0.15	0.15	0.15	0.16	0.16	0.16	0.16
B'A: $\omega 0$, $\omega 1$	0.17, 0.04	0.16, 0.12	0.16, 0.31	0.15, 0.69	0.17, 0.01	0.17, 0.04	0.16, 0.06	0.17, 0.02
Branch-site models								
A0: p2a ($\omega 2 = 1$)	0	0	0.08	0.18	0	0	0.91	0.10
AA: p2a, $\omega 2$	0, 1.00	0.04, 5.69	0.05, 7.79	0.12, 8.12	0, 1.00	0.01, 302.15	0.07, 9.67	0.13, 103.20
No. of filtered positive selected sites*	0	0	0	1	0	0	0	1
LRT, *P*								
B0 vs. BA	*<0.001*	*0.018*	*0.002*	0.063	*<0.001*	*<0.001*	*<0.001*	*<0.001*
B'0 vs. B'A	*<0.001*	*<0.001*	0.079	0.676	*<0.001*	*<0.001*	*<0.001*	*<0.001*
A0 vs. AA	0.994	*0.013*	*0.042*	*<0.001*	1.000	0.565	*0.004*	*<0.001*

Csol,*Ceratosolen solmsi*;*Coflo*, *Copidosoma floridanum*; *Mdem*, *Microplitis demolitor*; LRT, likelihood ratio test; ω, ratio of nonsynonymous to synonymous substitution rates; $\omega 0$, $\omega 1$, background and foreground ω values, respectively; *P*, LRT *P*-value (significant results in italics).
*Tests for positive selection on coding sequences for N-terminal (sex-determiner protein amino-terminal (SDP_N) domain and the autoregulation domain (CAM) domain) of *tra* homologues using the PAML branch model.
Selected parameter estimates and LRT results are shown.
The positive selection sites ($P > 0.95$) detected using branch-site model were then filtered to remove unreliable alignment columns using BALI-PHY's confidence scores (P1-max-AU). Number of positive selected sites ($P > 95\%$) with BALI-PHY's cofidence scores (P1-max-AU) >80% are shown.

species (Koch *et al*., 2014). In the ant *Harpegnathos saltator*, pseudogene-like sequences homologous to sequential *tra* exons are also present (*Hsal_traφ3* and *Hsal_traφ4*, Table S3). We found that the d_S divergences between *Hsal_traφ3* and *Hsaltra-A* and between *Hsal_traφ4* and *Hsaltra-A* were larger than those of the homologous regions of *Hsaltra-A* and *Hsaltra-B* (Table S4). This may be the result of faster evolution of pseudogenes compared with functional genes, or of both pseudogenes being duplicated prior to the split of the current functional gene copies of *Hsaltra-A* and *Hsaltra-B* (Fig. 2). These results indicate that independent whole-gene duplications of *tra* might have taken place in seven lineages (Fig. 1). It is also important to emphasize that independent duplications of *tra* are also present in the red harvester ant *Pogonomyrmex barbatus* and the fire ant *Solenopsis invicta*, which were not included in our study (Privman *et al*., 2013). These results concur with the prediction of both ourselves and of Koch *et al*. (2014) that *tra* genes and their pseudogenes have duplicated frequently and independently in each lineage of the Hymenoptera.

Evolutionary and expression analyses of the expanded tra *family in* Ce. solmsi

We found three members of the *tra* family in *Ce. solmsi*, the most in any species involved in the study. This is the first evidence for any *tra* duplication in any non-Aculeata hymenopteran species. As *Ce. solmsi* is an example of a non-CSD insect species, we thus conducted further investigations on its three *tra* members in an attempt to reveal their duplication history.

Selection on Csoltra-B *and* Csoltra-C *duplicates*. We measured the selective pressure on *tra* genes in chalcids by using the branch model in PAML (BA and B0 in Table 1). We found that although the ratios of nonsynonymous to synonymous substitution rates (ω) were usually less than 1, foreground ω values of *Csoltra-B* and *Csoltra-C* (0.28 and 0.30, respectively) were significantly higher than their background ω values (both 0.18). By contrast, foreground ω values were significantly smaller than background ω values for *Csoltra-A* and other Chalcidoidea *tra*. No significant fit was detected with a likelihood ratio test (LRT) for the *Csoltra-B/C* clade ($P = 0.063$, clade *f* in Fig. 1).

Because there are no models for insertions and deletions in the PAML programs, the abundant alignment gaps of *Csoltra-B* and *Csoltra-C* (Fig. S2) would consequently be treated as ambiguities, which may result in an underestimation of their sequence divergences (Yang, 2007). We therefore performed branch model analyses on the more conserved N-terminal domains (SDP_N domain and CAM domain; B'A and B'0 in Table 1). Our estimation of ω suggested that the majority of hymenopteran *tra* coding sequences are subject to strong functional constraints at the N-terminal ($P < 0.001$). However, we failed to obtain significantly improved fits in two clades according to a LRT, *Csoltra-C* and the common ancestor of *Csoltra-B* and *Csoltra-C* ($P = 0.079$ and 0.676, respectively).

We further performed branch-site model tests, which assume that ω may vary amongst both sites and branches. The results provided evidence for positive selection in at least one codon site in the SDP-N domain of *Csoltra-B/C* and one positively selected codon site in the Pro-rich domain of *M. demolitor tra* (AA and A0 in

Table 1). The results of the selection analysis in the branch test provided no evidence of positive selection, but the branch-site test showed the presence of at least one positively selected codon site in the two clades. A possible explanation for this may be that a mixture of many sites under purifying selection and few sites under positive selection results in low values of ω.

It is notable that in contrast with the two duplicated *tra* genes of *Ce. solmsi*, we detected no positive selection for *Csoltra-A*. We thus suggest that positive selection appears to be an important driver of the sequence divergence of *Csoltra-B* and *Csoltra-C*. It is possible that after duplication from the common ancestor, *Csoltra-A* maintains the original function of sex determination. However, the ancestor of *Csoltra-B* and *Csoltra-C* may be under positive selection.

Sex-specific alternative splicing (AS) and expression patterns of Csoltra *genes*. To learn more about the characteristics of the genes of the specific expanded *tra* family in *Ce. solmsi*, we analysed the expression of the *Csoltra* genes. For *Csoltra-A*, we detected three splice variants in females, named *Csoltra-A* F1–3 (Fig. 3A). Only *Csoltra-A* F1 could be translated into a full Fem protein with 399 amino acids (aa). In males four transcripts were produced but the predicted proteins were all truncated, generally because introns 4 and 5 were male-specifically spliced (Fig. 3A). Similar truncation of *tra* splice variants of males has been reported in other species (McKeown *et al*., 1987; Hasselmann *et al*., 2008; Werren *et al*., 2010; Saccone *et al*., 2011). We designed specific primers to amplify each of the transcripts (Fig. 3A, Table S5). All transcripts could be detected except for *Csoltra-A* F2 and F3, and *Csoltra-A* M4, three transcripts that may be rare splicing variants (Fig. 3D). Therefore, the main product of *Csoltra-A* in females was *Csoltra-A* F1, whereas males had at least three different transcriptional products (*Csoltra-A* M1–M3).

Neither *Csoltra-B* nor *Csoltra-C* had the same AS pattern as *Csoltra-A*. Cloning of the 3′ end of *Csoltra-B* resulted in three AS variants, *Csoltra-B* t1, t2 and t3 (Fig. 3B). However, we did not find AS in *Csoltra-C* (Fig. 3C). Our PCR amplification detected that all of the transcripts of *Csoltra-B* and *Csoltra-C* were specifically expressed in females, and that *Csoltra-B* t3 was the predominant transcript amongst the three *Csoltra-B* AS variants (Fig. 3E, F).

CsolTra protein domain arrangements as implied by the duplication history of Csoltra *genes*. To understand the divergence amongst the *Csoltra* paralogues, we aligned the protein sequence of *Csoltra-A* F1 and all the transcripts of the other two paralogues (Fig. S2). Amongst the three *Csoltra-B* transcripts, we found that CsolTra-B t3 and CsolTra-A were most similar to each other, whereas the AS 3′ ends of the other two transcripts (generated by exon skipping, Fig. 3B) varied considerably in their aa sequences. Considering that *Csoltra-B* t3 was the predominant transcript amongst the three AS forms, it might also be the only transcript that produces protein with a structure similar to that of CsolTra-A. The alignment (Fig. S2) showed that the predominant products of the three Tra proteins were conserved in the N terminus (SDP_N and CAM domains). By contrast, we detected evident sequence variations of both the SR and Pro-rich domains, especially for CsolTra-C, which had an extension of the SR domain in the C-terminal but lacked the Pro-rich domain.

The numerous Ser and Arg residues in the C-terminal of CsolTra-C initially appeared to us to form periodic groupings. Further analysis of repeats in the aa sequence revealed 17 repetitive Arg-Arg-Arg-Ser (RRRS) residues (Fig. S2). Furthermore, the comparison between residues 214–282 of CsolTra-B t3 and residues 242–520 of CsolTra-C showed a regular pattern of tandem 10-residue repeats in the aa sequence (SSSP/SRRRRSE/Q). CsolTra-B t3 had seven 10-residue repeats, which was treated as a 70-residue unit, and coincidentally, CsolTra-C had four such 70-residue units (Fig. 4). The five 70-residue units presented pairwise similarities ranging from 57.1 to 85.7% (Table S6). We suggest that a possible origin of the tandem repeat structure of CsolTra-C is an ancestral unit of 70 aa homologous to residues 214–282 of CsolTra-B. The duplicated units may have evolved by unequal crossing over and subsequent gene conversion.

Discussion

Although the sex-determination gene *tra* in the Hymenoptera has been frequently and recently investigated, the present study provides novel data for hymenopteran *tra*, enabling the construction of more robust phylogenies. After searching the genome assemblies and predicted gene sets several species, we found three *tra* copies in the fig wasp *Ce. solmsi*, the most expanded *tra* family. Considering that the Symphyta, the basal lineage of Hymenoptera, possess only a single copy of *tra*, we suggest that our new data provide novel evidence in support of multiple, independent *tra* duplications in the Hymenoptera. The sporadic distribution of *tra* duplicates amongst different hymenopteran lineages supports the hypothesis of multiple independent duplications rather than an ancestral duplication followed by concerted evolution (Schmieder *et al*., 2012; Privman *et al*., 2013; Koch *et al*., 2014).

Our analyses of hymenopteran *tra* pseudogenes also show a pattern consistent with the birth-and-death model of evolution. In this model, new genes are created by duplication and some of these duplicated genes are deleted or become nonfunctional by deleterious

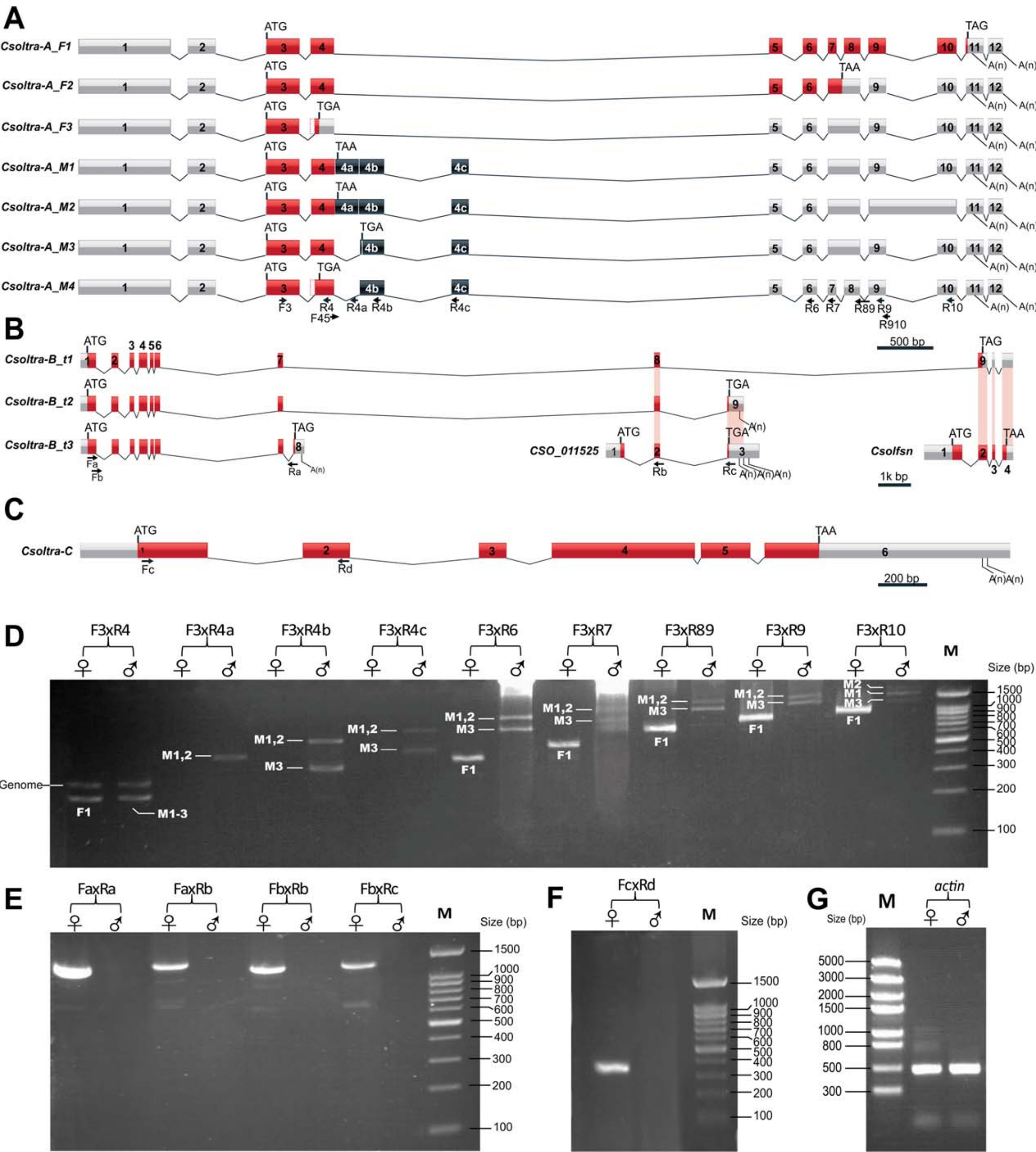

Figure 3. Alternative splicing (AS) of *Ceratosolen solmsi transformer* (*Csoltra*) genes. (A–C) Exon and intron structure diagrams of all the transcripts of the three *Csoltra* genes showing the primer positions. Exons are shown as boxes and introns as lines. Predicted coding sequences are marked in red, male-specific exons of *Csoltra-A* are in black and untranslated region exons (UTRs) in grey. Polyadenylation sites are shown as A(n)s. Horizontal arrows show the locations of the specific primers used to verify the transcripts in RT-PCR experiments. (D–F) Identification of *Csoltra* transcripts by RT-PCR in female (♀) and male (♂) adults. (D) RT-PCR of *Csoltra-A* using the combinations of one common forward primer (F3) and nine different reverse (R4–R10) primers (locations are shown in A). The detected male-specific *Csoltra-A* AS products, *Csoltra-A* M1, M2 and M3, are shown as white lines; only one female-specific AS form was detected (*Csoltra-A* F1). Owing to the short products of F3 × R4 primer combination, the by-products of genomic sequences were also amplified during PCR. (E) RT-PCR of *Csoltra-B* using the primer combinations Fa × Ra, Fa × Rb, Fb × Rb and Fb × Rc (primer locations are in B). (F) reverse transcription PCR (RT-PCR) products of *Csoltra-C* generated by the primer combination Fc × Rd (primer locations are in C). (G) *Csolactin* gene was amplified from the same samples to examine the quality of cDNAs. Size markers (M) are indicated in base pairs. Primer sequences used are presented in Table S5.

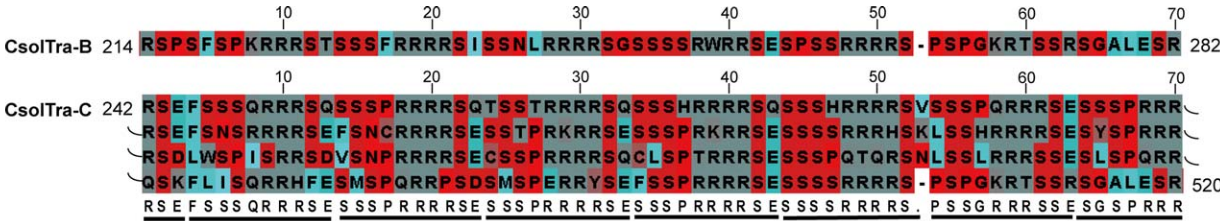

Figure 4. Alignment of amino acid sequences of residues 214–282 of *Ceratosolen solmsi* Transformer-B (CsolTra-B) and residues 242–520 of CsolTra-C. Residues 214–282 of CsolTra-B and 242−520 of CsolTra-C are aligned to show the repeated units. The best-conserved residues are shown below the alignment. The black line shows the repeated SSSP/SRRRRSE/Q unit. The sequence alignment is viewed and coloured in Jalview (Waterhouse *et al.*, 2009).

mutations, whereas others may remain in the genome for a long time (Hughes & Nei, 1989). The abundant *tra* pseudogenes that we found in the Formicidae and in the Apoidea suggest that duplication events deriving novel *tra* genes or pseudogenes may be common in the Hymenoptera, especially in ants and bees. By contrast, most of the species of Chalcidoidea, Braconidae and Symphyta examined here did not show evidence of possessing paralogous *tra* (except *Ce. solmsi*) or pseudogenes. Our new discoveries of lineage-specific distributions of *tra* pseudogenes, which may be the remnants of gene duplications, suggest repeated independent duplications of *tra* in the Hymenoptera, especially in the Aculeata. This is because frequent gene duplications and losses involved in the birth-and-death evolution process always generate pseudogene fragments scattering around the functional copies.

No *tra* gene duplication event has been found outside the Aculeata before (Geuverink & Beukeboom, 2014). In our search for *tra* homologues and pseudogenes in three non-CSD Chalcidoidea genomes (*Ce. solmsi*, *T. pretiosum* and *Co. floridanum*), we found three copies of *tra* in *Ce. solmsi*, with only a single copy of *tra* present in the other two species. Our discovery of repeatedly duplicated *tra* copies in *Ce. solmsi* is not only consistent with the hypothesis of independent duplication for hymenopteran *tra*, but also provides the first example of *tra* evolution amongst non-CSD hymenopterans.

By further tracing the duplication trajectory of the *tra* gene using evolutionary and expression analyses, we found that *Csoltra-B* and *Csoltra-C* may be duplicated from *Csoltra-A* and encode novel SR-related proteins. Gene duplications are common in SR protein evolution, with the expansion of SR repeats being widely thought to play a crucial role in the evolution of splicing regulation and in the relaxation of splicing signals (Shepard & Hertel, 2009). Previous studies have suggested that SR-rich domains regulate splicing by protein–protein interactions, and that Pro-rich domains also have a protein-binding function (Graveley, 2000; Kay *et al.*, 2000; Hastings & Krainer, 2001). Variations of both SR-rich and Pro-rich domains in these *Csoltra* paralogues suggest variable regulation functions of the *tra* family in *Ce. solmsi*. Further investigation is necessary to determine if the extended SR domain of *Csoltra-C* and the variable SR domain of *Csoltra-B* offer binding sites for associated proteins to regulate gene splicing. Our selection analysis provided evidence that the common ancestor of *Csoltra-B* and *Csoltra-C* was subject to positive selection (Table 1). We therefore conclude that positive selection is instrumental in the differentiation of *Csoltra* paralogues. Further expression analyses as shown in our reverse transcription PCR (RT-PCR) results show that the expression of both paralogues is female-specific, indicating possible sex-related functions. Although we cannot yet conclude that both genes are involved in the process of sex determination in this fig wasp species, this sex-specific expression refutes the possibility of the pseudogenization of *Csoltra-B* or *Csoltra-C*.

Small-scale mutational changes in the nucleotide sequences of duplicated copies of genes can generate novel sex-determination mechanisms (Gempe & Beye, 2011; Bopp *et al.*, 2014). The evolution of duplicated *tra* genes in a non-CSD hymenopteran species enables us to hypothesize that the recent duplication of *Csoltra-B* and *Csoltra-C* may have favoured the origination of a novel sex-determination pathway in the fig wasp. It has been proposed that gene duplication and positive selection can shape the growth of a new upstream signal and create a novel sex-determination pathway, such as *csd* in *A. mellifera* (Hasselmann *et al.*, 2008). We also found a signature of positive selection for *Csoltra-B/C*, suggesting that positive evolution is involved in the process of neofunctionalization of these new paralogues. Additionally, there are other examples of nucleotide changes in duplicated genes generating novel sex-determination genes and new sex-determination mechanisms, such as the evolutionary rise of *csd* in *Apis* and *sxl* in *D. melanogaster* (Traut *et al.*, 2006; Hasselmann *et al.*, 2008). Previous studies have also shown that the sexually spliced *tra* of a recently evolved allele can dominate the

sex-determination pathway in *Musca domestica* (Hediger *et al.*, 2010). *Csoltra-B* and *Csoltra-C* are female-specific in their expression, and are transcribed without any sex-specific splicing. This is the first report of sexually dimorphic expression of *tra* duplicates, which suggests some possible sex-related functions of the two paralogues. Although further investigations are needed, the unexpected presence of the two paralogues *Csoltra-B* and *Csoltra-C* makes the non-CSD fig wasp *Ce. solmsi* a remarkable hymenopteran example that is likely to broaden our understanding of sex determination and the evolution of new gene function in general.

Experimental procedures

Genome searches for tra *homologues and pseudogenes*

We identified homologous *tra* sequences in five hymenopteran genome assemblies that had all been deposited in the i5k Genome Database (http://www.arthropodgenomes.org/wiki/i5K). We performed local BLAST searches (Altschul *et al.*, 1990) using previously published protein sequences of Tra from closely related species and predicted gene structures with FGENESH+ in the SOFTBERRY software package (http://linux1.softberry.com; Solovyev, 2001). Exon boundaries were manually adjusted with the software package IGV (Thorvaldsdottir *et al.*, 2013). The homologous *tra* genes used in this study are shown in Table S2. We searched for pseudogenes using BLASTN (Altschul *et al.*, 1990); annotated or published *tra* genes of each hymenopteran species being examined were used as queries. Genomic sequences with significant similarity (with E-values $< 10^{-5}$ as a cut-off) but without full coding frames were treated as pseudogenes. We selected pseudogene segments located in the same scaffold and showing the highest similarity to the same *tra* paralogue. If they had the same genomic orientations, and had sequence similarity to one *tra* exon or several sequential ordered *tra* exons, they were treated as a single degenerated pseudogene. Pseudogenes that failed to fulfil these criteria were numbered separately. All pseudogenes were numbered independently for each species and so the pseudogene numbers as listed in Table S3, S4 and Fig. 2 do not represent any homologous relationship of the pseudogenes.

Phylogenetic and molecular evolutionary analyses

We used BALI-PHY, a program that minimizes the impact of length variation of sequences by estimating the alignment and tree topology simultaneously using a Markov Chain Monte Carlo sampler (Suchard & Redelings, 2006), to align the sequences. The inferred aa sequences were initially aligned using MAFFT (Katoh & Standley, 2013) to enable subsequent estimation of the best-fitting aa substitution model with PROTTEST (Abascal *et al.*, 2005). The Jones-Taylor-Thornton (JTT) model with gamma distribution of rates and invariant site categories (inv) was set for the final alignment using BALI-PHY, and for the phylogeny inferences using BALY-PHY and PHYML3.0 (Guindon *et al.*, 2010). The reversible Markov model of mitochondrial amino acids (Mtrev) + gamma + inv model was used in the Bayesian phylogeny inference using MRBAYES (Huelsenbeck & Ronquist, 2001). For the BALI-PHY analyses, six independent runs were performed. For each run 100 000 iterations were executed. The first 10 000 samples of each run were discarded as burn-in, with a majority-rule consensus tree being obtained from all runs after burn-in.

We used the CODEML program within PAML 4.5 (Yang, 1997) to evaluate the role of natural selection during the evolution of *tra* genes in Chalcidoidea. The d_N/d_S ratios (d_N = nonsynonymous per site substitutions) were estimated to test the potential positive selection, using branch model and branch-site model A. Several lineages of chalcidoids were independently labelled, and we ran once for each of them as the foreground branch. The branch model (two-ratio) was tested ($df = 1$) against a null model (fixed ω of 1). The branch-site model (model = 2, NSsites = 2) was also tested ($df = 1$) against a null model (fixed ω of 1). For each examined branch, we calculated the LRT comparing the paired models. Positively selected sites were identified using Bayes empirical Bayes criteria for each branch-site model (Yang *et al.*, 2005). The positive selection sites ($P > 0.95$) were then filtered to remove unreliable alignment columns using BALI-PHY's confidence scores [(P1-max-AU); Suchard & Redelings, 2006]. Only positive selection sites ($P > 0.95$) with P1-max-AU values > 0.80 were concluded to be under positive selection.

Cloning and transcription analyses of tra *and its paralogues in the* Ce. solmsi

Adults of *Ce. solmsi*, the pollinator of *Ficus hispida*, were collected from the figs of naturally growing trees in Danzhou, Hainan province, China (19°30′N, 109°31′E), in July 2012. Total RNA from a single individual of both gender was isolated using EasyPure^{TM} RNA kits (TransGen Biotech, Beijing, China) following the manufacturer's instructions. First-strand cDNA was then synthesized from 500 ng of total RNAs using TransScript II First-Strand cDNA Synthesisi SuperMix (TransGen Biotech), with a random Oligo $(dT)_{20}$ primer according to the recommendations of the manufacturer. 5′- and 3′-rapid amplification of cDNA ends (RACE) was performed using a BD SMART^{TM} RACE cDNA Amplification kit (BD Biosciences Clontech, San Jose, CA, USA). The primers used for RACE and sex-specific transcription analyses are presented in Table S5.

Acknowledgements

This project was supported by the National Natural Science Foundation of China (NSFC grant nos 31210103912, 31422050), partially by a grant (O529YX5105) from the Key Laboratory of Zoological Systematics and Evolution of the Chinese Academy of Sciences, and the National Science Fund for Fostering Talents in Basic Research (Special Subjects in Animal Taxonomy, NSFC-J0930004). We thank Dr Wen Xin and TransGen Biotech for providing some of the reagents used for the study. We thank the anonymous reviewers for their valuable comments and suggestions. We thank Derek W. Dunn for suggestions that improved this paper.

References

Abascal, F., Zardoya, R. and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**: 2104–2105.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Asplen, M.K., Whitfield, J.B., DE Boer, J.G. and Heimpel, G.E. (2009) Ancestral state reconstruction analysis of Hymenopteran sex determination mechanisms. *J Evol Biol* **22**: 1762–1769.

Belote, J.M., McKeown, M., Boggs, R.T., Ohkawa, R. and Sosnowski, B.A. (1989) Molecular genetics of *transformer*, a genetic switch controlling sexual differentiation in *Drosophila*. *Dev Genet* **10**: 143–154.

Beukeboom, L.W. (1995) Sex determination in Hymenoptera: a need for genetic and molecular studies. *Bioessays* **17**: 813–817.

Beukeboom, L.W., Kamping, A. and van de Zande, L. (2007) Sex determination in the haplodiploid wasp *Nasonia vitripennis* (Hymenoptera: Chalcidoidea): a critical consideration of models and evidence. *Semin Cell Dev Biol* **18**: 371–378.

Beye, M., Hasselmann, M., Fondrk, M.K., Page, R.E. and Omholt, S.W. (2003) The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. *Cell* **114**: 419–429.

Bopp, D., Saccone, G. and Beye, M. (2014) Sex determination in insects: variations on a common theme. *Sex Dev* **8**: 20–28.

Danforth, B.N., Cardinal, S., Praz, C., Almeida, E.A.B. and Michez, D. (2013) The impact of molecular data on our understanding of bee phylogeny and evolution. *Annu Rev Entomol* **58**: 57–78.

Gempe, T. and Beye, M. (2011) Function and evolution of sex determination mechanisms, genes and pathways in insects. *Bioessays* **33**: 52–60.

Geuverink, E. and Beukeboom, L.W. (2014) Phylogenetic distribution and evolutionary dynamics of the sex determination genes *doublesex* and *transformer* in insects. *Sex Dev* **8**: 38–49.

Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA* **6**: 1197–1211.

Grbic, M., Ode, P.J. and Strand, M.R. (1992) Sibling rivalry and brood sex-ratios in polyembryonic wasps. *Nature* **360**: 254–256.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.

Hasselmann, M., Gempe, T., Schiott, M., Nunes-Silva, C.G., Otte, M. and Beye, M. (2008) Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* **454**: 519–522.

Hastings, M.L. and Krainer, A.R. (2001) Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* **13**: 302–309.

Hediger, M., Henggeler, C., Meier, N., Perez, R., Saccone, G. and Bopp, D. (2010) Molecular characterization of the key switch F provides a basis for understanding the rapid divergence of the sex-determining pathway in the housefly. *Genetics* **184**: 155–170.

Hedley, M.L. and Maniatis, T. (1991) Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to *tra-2* protein in vitro. *Cell* **65**: 579–586.

Heraty, J.M., Burks, R.A., Cruaud, A., Gibson, G.A.P., Liljeblad, J., Munro, J. *et al*. (2013) A phylogenetic analysis of the megadiverse Chalcidoidea (Hymenoptera). *Cladistics* **29**: 466–542.

Huelsenbeck, J.P. and Ronquist, F. (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**: 754–755.

Hughes, A.L. and Nei, M. (1989) Evolution of the major histocompatibility complex - independent origin of nonclassical class-I genes in different groups of mammals. *Mol Biol Evol* **6**: 559–579.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772–780.

Kay, B.K., Williamson, M.P. and Sudol, P. (2000) The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14**: 231–241.

Koch, V., Nissen, I., Schmitt, B.D. and Beye, M. (2014) Independent evolutionary origin of *fem* paralogous genes and complementary sex determination in Hymenopteran insects. *PLoS ONE* **9**: e91883.

Kukuk, P.F. and May, B. (1990) Diploid males in a primitively eusocial bee, *Lasioglossum* (*Dialictus*) *Zephyrum* (Hymenoptera, Halictidae). *Evolution* **44**: 1522–1528.

Lynch, K.W. and Maniatis, T. (1996) Assembly of specific SR protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer. *Genes Dev* **10**: 2089–2101.

Mao, M., Gibson, T. and Dowton, M. (2015) Higher-level phylogeny of the Hymenoptera inferred from mitochondrial genomes. *Mol Phylogenet Evol* **84**: 34–43.

McKeown, M., Belote, J.M. and Baker, B.S. (1987) A molecular analysis of *transformer*, a gene in *Drosophila melanogaster* that controls female sexual differentiation. *Cell* **48**: 489–499.

McKeown, M., Belote, J.M. and Boggs, R.T. (1988) Ectopic expression of the female *transformer* gene product leads to female differentiation of chromosomally male *Drosophila*. *Cell* **53**: 887–895.

Moreau, C.S. and Bell, C.D. (2013) Testing the museum versus cradle tropical biological diversity hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution* **67**: 2240–2257.

Peng, Y.Q., Zhang, Y., Compton, S.G. and Yang, D.R. (2014) Fig wasps from the centre of figs have more chances to mate, more offspring and more female-biased offspring sex ratios. *Anim Behav* **98**: 19–25.

Privman, E., Wurm, Y. and Keller, L. (2013) Duplication and concerted evolution in a master sex determiner under balancing selection. *Proc Biol Sci* **280**: 20122968.

Saccone, G., Salvemini, M. and Polito, L.C. (2011) The *transformer* gene of *Ceratitis capitata*: a paradigm for a conserved epigenetic master regulator of sex determination in insects. *Genetica* **139**: 99–111.

Schmieder, S., Colinet, D. and Poirie, M. (2012) Tracing back the nascence of a new sex-determination pathway to the ancestor of bees and ants. *Nat Commun* **3**: 895.

Shepard, P.J. and Hertel, K.J. (2009) The SR protein family. *Genome Biol* **10**: 242.

Smith, B.H. (1983) Recognition of female kin by male bees through olfactory signals. *Proc Natl Acad Sci USA* **80**: 4551–4553.

Solovyev, V. (2001) Statistical approaches in eukaryotic gene prediction. In *Handbook of Statistical Genetics* (Balding, D. et al., eds), pp. 83–127. John Wiley & Sons, Ltd.

Stouthamer, R. and Kazmer, D.J. (1994) Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in *Trichogramma* wasps. *Heredity* **73**: 317–327.

Suchard, M.A. and Redelings, B.D. (2006) BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**: 2047–2048.

Thorvaldsdottir, H., Robinson, J.T. and Mesirov, J.P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192.

Traut, W., Mimi, T., Ikeo, K. and Sahara, K. (2006) Phylogeny of the sex-determining gene *sex-lethal* in insects. *Genome* **49**: 254–262.

Verhulst, E.C., van de Zande, L. and Beukeboom, L.W. (2010) Insect sex determination: it all evolves around *transformer*. *Curr Opin Genet Dev* **20**: 376–383.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K. *et al*. (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science* **327**: 343–348.

Whiting, P.W. (1939) Sex determination and reproductive economy in *Habrobracon Genetics* **24**: 111.

Whiting, P.W. (1943) Multiple alleles in complementary sex determination of *Habrobracon*. *Genetics* **28**: 365.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555–556.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**: 1586–1591.

Yang, Z., Wong, W.S. and Nielsen, R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**: 1107–1118.

van de Zande, L. and Verhulst, E.C. (2014) Genomic imprinting and maternal effect genes in haplodiploid sex determination. *Sex Dev* **8**: 74–82.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. Transformer (Tra) phylogeny as determined using Bayesian inference in MrBayes and maximum-likelihood analyses in PhyML. (A) Bayesian phylogenetic analysis using MrBayes running for 600 000 generations (162 000 burn-in). Starting trees were random and trees were sampled every 100th generation. Posterior probability support is indicated for each node. (B) Phylogeny analysis using PhyML on 1000 replicates. The scale bar represents the estimated number of substitutions per site. Bootstrap proportions are shown above each branch. Gene abbreviations are as shown in Fig. 1 for both trees.

Figure S2. Transformer (Tra) protein domain arrangements in *Ceratosolen solmsi* (*Csol*). Aligned predicted protein sequences of *Csoltra* homologues. The functional transcripts of *Csoltra-A* F1, all three alternative splicing transcripts of *Csoltra-B*, and *Csoltra-C* are aligned. The black lines show the 17 RRRS repeats in *Csoltra-C*. Multiple sequence alignments were generated by Bali-Phy and adjusted manually.

Table S1. *Transformer* (*tra*) homologues annotated in the newly available hymenopteran genomes.

Table S2. Published *transformer* (*tra*) homologues used for this study.

Table S3. *Transformer* (*tra*) pseudogenes in the Hymenoptera.

Table S4. Comparison of the synonymous per site substitutions (*ds*) between each pseudogene and its parent gene, and the *ds* between the most related *transformer* (*tra*) gene pairs.

Table S5. Primers for sex-specific transcripts of Ceratosolen solmsi transformer (*Csoltra*) homologues genes.

Table S6. Pairwise amino acid similarities between 70-residue units for *Ceratosolen* Transformer-B (CsolTra-B) and C.