Box 7.1 The identification of *tra2*, *tra* and *dsx* and their ever-transforming nomenclature

The strong sequence divergence of *tra* and *dsx* leads to ambiguities in their identification and nomenclature. Duplications of *tra* complicate the identification of sex determination genes further. Newly identified genes in the sex determination cascade are often given similar names, due to their feminizing, masculinizing and transforming nature. This box discusses confusing patterns of sex determination gene nomenclature and explains how I have dealt with paralogs and putative absences of genes in this thesis.

Recognition of tra2

Tra2 is remarkably conserved compared to *dsx* and *tra* and is recognized by its RNA-binding domain (RBD). Despite their matching names, *tra* and *tra2* have no known shared origin and their nomenclature is based on their roles in *Drosophila* sex determination. The high conservation of *tra2* could be related to its functions in early development other than sex determination, as it is involved in spermatogenesis in *D. melanogaster* (Hazelrigg & Tu, 1994; Madigan *et al.*, 1996; Mattox *et al.*, 1996) and embryogenesis in *A. mellifera* (Nissen *et al.*, 2012) and *N. vitripennis* (Chapter 3). Furthermore, the role of *tra2* in sex determination is functionality conserved in species seemingly lacking a *tra* ortholog (Suzuki *et al.*, 2012). *Tra2* and *tra* interact to form a protein-complex in *D. melanogaster* (Amrein *et al.*, 1994), which binds to a regulatory element on the *dsx* gene (Hedley & Maniatis, 1991; Inoue *et al.*, 1992; Tian & Maniatis, 1992). It has not been assessed whether *tra2* orthologs have a higher conservation in species that also carry *tra*, or contain particular sequence elements, compared to species that apparently lack *tra*.

Recognition of dsx

Dsx is characterized by its combination of a N-terminus DNA-binding motif dsx/mab3 (DM) domain with a oligomerization domain (OD1) and a second C-terminus oligomerization (dimer) domain (OD2) (reviewed in Verhulst & van de Zande (2015)). The DM domain is not unique to *dsx* and is also found in other genes. Reconstructing their evolutionary relationship is complicated by the lack of sequence similarity apart from the DM domain. Collectively, they form the class of *Dmrt* transcription factors that are present throughout Metazoans and play a role in establishing sexual dimorphisms (Kopp, 2012). The distinction between *dsx* and other *Dmrt* genes is based on the second dimer domain that is unique to *dsx* and characterized by divergent sex-specific isoforms. Yet, the absence of a OD2 domain can be overlooked and sex-specific splicing is either not tested or not viewed as an identifier of *dsx*. This has led to some unreliable characterizations of *dsx*, the most persistent of which is the *Daphnia dsx* conundrum. *Daphnia* possess two genes which resemble *dsx* (Kato *et al.*, 2011; Toyota *et al.*, 2013), but the alignments of its potential OD2 domain do not show conservation alike the strong similarity between insect DSX isoforms. Furthermore, neither copy displays sex-specific splicing. The DM-

domain genes of these crustaceans appear as a separate clade in gene trees of *Dmrt* homologs (Wexler *et al.*, 2014). The assignment of *dsx* may need to be restricted to insects and requires the presence of sex-specific splicing and, consequently, the presence of sex-specific OD2 domains.

Recognition of tra

Whereas *dsx* can at least be recognized by two specific regions (DM and OD), this unfortunately cannot be said for *tra*. The CAM domain is found in all studied insect species containing a *tra* ortholog, except *Drosophila*, thus allowing some discriminatory classification (Verhulst *et al.*, 2010b). The *Drosophila* exception straight away complicates matters, as its CAM domain absence appears to result from the addition of an upstream element of autoregulation (*sxl*) to the sex determination cascade (Bopp, 2010). This particular addition has not been found outside of *Drosophila*, but with the tendency of the sex determination cascade to evolve upwards (Wilkins, 1995) in mind, the addition or incorporation of new upstream elements may occur in a variety of ways. *Drosophila tra* has the characteristic Diptera-specific (DIP) domain that has subsequently been identified in all dipteran species thus far. Even the dipteran, *Phlebotomus papatasi,* whose candidate *tra* appears to lack the CAM domain (Chapter 2), putatively possesses the DIP domain. Hymenoptera have another order-specific domain (HYM), which does not display any similarity in amino acid sequence with its dipteran counterpart. The function of both the HYM domain as well is the DIP domain is unknown. The HYM domain is present in all hymenopteran *tra* homologs, irrespective of whether these homologs have an apparent orthologous function or a potential paralogous function from duplicated origin. In all insects, two regions are common in *tra*: an arginine-serine and a proline-rich region. However, these motifs are also abundantly present in other genes and, as such, do not aid in *tra* identification without additional information about the order-specific domain or the CAM domain.

Duplications of tra

Paralogous *tra* sequences can be highly conserved, as is for example the case for the duplicates *tra* and *traB* in *L. clavipes* (Chapter 5). The presence of sex-specific splicing can be the first discriminating factor to assign a putative orthologous relationship. *Tra* in *L. clavipes* is spliced into a male and a female mode, whereas *traB* does not undergo sex-specific splicing. A similar pattern is found in the *fem* and *csd* pair of *A. mellifera*, where *fem* has sex-specific splicing, yet *csd,* despite containing all *tra* characteristic domains, displays only minor differences between its two forms (one predominant and one rare) (Beye *et al.*, 2003; Hasselmann *et al.*, 2008a). *Csd* genes which possess an actual CSD function also code for a specific hypervariable region that contains asparagine- and tyrosine-enriched repeats, which are responsible for the high allelic variation at this locus (Hasselmann *et al.*, 2008b). Thus far no duplicates of *tra* have been found that exhibit sex-specific splicing and the hypervariable *csd* region has not been documented outside *Apis* species. Functional studies of *tra* duplicates are needed to determine their

potential role in sex determination and the evolutionary history of sex determination genes in the different branches of Hymenoptera.

Gene-pairs of putative *tra* orthologs and their paralogous copies have thus far been named *traA*/*traB* (Wurm *et al.*, 2011; Privman *et al.*, 2013), *fem*/*csd* (Schmieder *et al.*, 2012), and *fem*/*fem1* (Koch *et al.*, 2014). The nomenclature of *fem* and *csd* is characteristic for *A. mellifera* and closely related species. In these systems the *csd* gene performs the complementary function for sex determination. Applying this nomenclature in ants and other less related species seems premature, as there is no indication that the *traB*/*csd*/*fem1* paralog actually fulfills a *csd* function. An argument against *csd* functionality is that the hypervariable region of *Apis csd* is consistently absent in these *tra* duplicates (Koch *et al.*, 2014). The nomenclature of *fem* in Hymenoptera may be more fitting, due to apparent conservation of its feminizing function in this order. It may however be wise to limit this to CSD systems, or change its name to *tra* for the entire Hymenoptera, as the *Drosophila* designation of *tra* predates *fem*. *Fem* nomenclature was recently further complicated by the discovery of a feminizing gene in *Bombyx mori*, which was inconveniently named *feminizer* (Kiuchi *et al.*, 2014). The *fem*/*fem1* use, suggested by (Koch *et al.*, 2014), may thus not be the best option if functionality has not been determined and other insect sex determination genes are called *fem* as well. The plurality of genes called *fem* has already resulted in publications of the wrong *fem* (i.e. *fem-1* of *Caenorhabditis elegans*) in *Diachasmimorpha longicaudata* (Mannino *et al.*, 2016). This leaves the *traA*/*traB* (and *traC* (Jia *et al.*, 2016)) nomenclature as the best option in cases where *tra* functionality is not yet known. Paralogs of *tra* ought to be separated by alphabetical markings, as numerical distinctions will overlap with *tra2*. In studies that have evidence for which copy is orthologous to *tra* of other insects, this gene-pair may be called *tra*/*traB*, as we did in Chapter 5 for *L. clavipes*.

Absence of tra

While the identification of multiple *tra* copies has its technical difficulties, a possible lack of *tra* is even harder to pinpoint. Some homology with previously described *tra* sequences is necessary to infer new orthologs, as all canonical *tra* share both CAM or order-specific domains, and a sex determination function through its interaction with *tra2*. A gene with similar functionality, but no traceable ancestry to annotated *tra* orthologs requires a different name, but can be grouped in the *tra*-like gene family.

Suggested absences of *tra* have thus far been reported for Lepidoptera and species in the Nematocera lineage of the Diptera (Mita *et al.*, 2004; Salvemini *et al.*, 2013) (Chapter 2). These groups have a large evolutionary distance to study systems in which *tra* has been identified. There are only two lepidopteran genera whose sex determination system has been investigated in any detail, the silk moth *B. mori* and *Ostrinia* moths (discussed in the main text of this chapter). If *tra* is present in Lepidoptera, it may contain a yet unknown Lepidoptera-specific domain. As no CAM-domains have been detected in Lepidoptera, a possible *tra*-like candidate may only be identified from its potential interaction with *tra2*. *Tra2* in *B. mori* does not control

the sex-specific splicing of *dsx* pre-mRNA (Suzuki *et al.*, 2012), an important distinction from the function of *tra2* in Diptera and Hymenoptera. Thus, evolution of the sex determination cascade of *B. mori*, and maybe of the Lepidoptera as a whole, may have taken an entirely different route at the TRA/TRA2 transduction level, potentially leading to a loss of *tra*.

The second possible case of *tra* absence occurs in branches of the Nematocera, the dipteran clade that comprises mosquitoes. Extensive genomic and transcriptomic data are available for genera such as *Anopheles*, *Aedes* and *Culex*. Yet, no homology with other dipteran *tra* can be detected (Chapter 2). Inferring conservation of the sex determination cascade is trickier in Nematocera, as no studies have been performed on *tra2*. These systems still contain a black box between the primary signal/upstream factors and the conserved *dsx* at the bottom of the cascade.

Tra may be recruited in the sex determination cascade at the start of the holometabolous insects (Chapter 2). Homologs of *tra* or traces of CAM domains appear in some groups outside the holometabolous insects, often at great evolutionary distance. Two systems have been studied in detail: *tra* expression during development in *Daphnia* (Kato *et al.*, 2010; Chen *et al.*, 2014; Kong *et al.*, 2015) and the functional analysis of a *tra* homolog in the acorn worm *S. kowalevskii* (Suzuki *et al.*, 2012). Unlike the ambiguity in the *Daphnia dsx* sequences, these putative *tra* genes do possess a CAM domain and arginine-serine regions. A lack of sex-specific *tra* splicing, and no sex reversal after knockdown of *tra* in *S. kowalevskii*, suggests that these genes are not part of sex determination in these species. These *tra-like* genes may form a distinct clade, alike the *Daphnia dsx-like* clade described above, which may need to be reflected in the nomenclature.

The Hymenoptera provide an exciting possibility to study evolutionary loss (and gain) of *tra*. *Tra* homologs, often multiple per species across families (unpublished data), can be detected relatively easily in Hymenoptera due to their conserved combination of the HYM domain and the CAM domain. This pattern of *tra* presence appears to hold for all branches of Hymenoptera, except one sawfly superfamily. The transcriptome of *Tenthredo koehleri* (Misof *et al.*, 2014) does not contain *tra* homologs and this pattern extends to other Tenthredoidea (unpublished data). *Tra2* is conserved in these species. Evolutionary distances are large, 250 MYA, but the ubiquity of *tra* in other hymenopteran lineages (unpublished data) and the appearance of CSD in this superfamily (van Wilgenburg *et al.*, 2006) make the tenthredoidid branch of great interest for further study into the evolutionary history of *tra*.

Additional functions of *tra* and *tra2*

Whereas the insect sex determination pathway is described as a linear cascade, more interacting components have been identified in vertebrate systems, resulting in elaborate gene-regulatory networks (Herpin & Schartl, 2015). Sex-specific splicing regulation appears to be a hallmark of insect sex determination. The sex-restricted splice forms of sex determination genes might limit additional functionality in developmental programs. *Tra2* is notably not sex-specifically spliced and indeed appears to have other functions in embryogenesis of *A. mellifera* (Nissen *et al.*, 2012) and *N. vitripennis* (Chapter 3). The abnormal alternative splicing of *tra2* in *A. japonica* (Chapter 6), strikingly absent in its close relative *A. tabida* (Chapter 4), could relate to this function in embryogenesis as the longer splice forms disappear during larval development. Preliminary results, however, indicate little or no effect of *tra2* pRNAi knockdown on embryogenic development in *A. japonica*, whereas its critical role as maternally provided sex determination gene has remained intact (unpublished data). Oogenesis did not appear impaired and suggests that *tra2* in *A. japonica* does not have a similar function in spermatogenesis of *tra2* in *D. melanogaster*, the single other case of sex-specific *tra2* splicing (Mattox & Baker, 1991; McGuffin *et al.*, 1998).

The only known targets of *tra* are the transcription factors *dsx* and *fruitless* (*fru*) (Hoshijima *et al.*, 1991; Inoue *et al.*, 1992; Heinrichs *et al.*, 1998) and until recently no functionality of *tra* apart from downstream sexual differentiation had been detected. Rideout *et al.* (2015) demonstrated effects of *tra* on *D. melanogaster* body size independently of *dsx* and *fru*, indicating a different branch in the *tra* regulatory network. It is not yet known whether this function is restricted to *Drosophila* or whether it is a conserved feature of more insect groups. The strong sequence divergence of *tra* may, however, also indicate a limited conservation of potential ancillary functions. Hints at additional *tra* functions do exist, e.g. in diapause induction of *tra*, but not *tra2*, after pRNAi in *N. vitripennis* (unpublished data). Further study of the regulatory functions of *tra* is clearly needed.

The diploid mortality upon *tra* and *tra2* knockdown in *N. vitripennis* (Chapter 3) may at first sight indicate additional *tra* and *tra2* functionality. Reduction of *tra* mRNA levels however did not impact haploid mortality. *Tra2* seems to have two separate mortality effects: one independent of *tra* in embryogenesis, and a second causing diploid mortality alike the one observed for *tra* knockdown. Many questions about the roles of *tra* and *tra2* in development remain open, and this is a promising area for future research in evo-devo.

Sex determination in *Asobara* and *Leptopilina*

CSD has been excluded for *Asobara* species (Beukeboom *et al.*, 2000; Ma *et al.*, 2013) and deemed highly unlikely in *L. clavipes* based on a lack of male-biased sex ratios in highly inbred lineages (K. Kraaijeveld, pers. comm.). Our results obtained for *A. tabida*, *A. japonica* and *L. clavipes* may fit the MEGISD model, albeit with modifications, as these species lack a number of discriminating features compared to *Nasonia*.

First, *tra*^{F} maternal provision is not required for female development of diploid fertilized eggs. Other maternally provided factors may fulfill this role, but are likely distinct from the autoregulatory function of *tra*^{F}. RNAi studies should shed light on the maternal effects involved in *Asobara* and *Leptopilina* sex determination. If *tra* knockdown in arrhenotokous females would lead to diploid male development it could indicate that *tra*^{NSS} is involved in sex determination. Furthermore, it would group this potential mechanism as one including a maternal effect component, wherein the action of this maternal effect is switched from maternal provision of *tra*^{F} mRNA to alternative forms.

Second, in light of the cascade evolving upward, it appears unlikely that the primary signals will be shared between *Asobara*, *Leptopilina* and *Nasonia*. *A. tabida* and *A. japonica* may use the same primary signal, though the lack of *tra* sequence conservation may hold surprises. Either a fertilization factor (e.g. a short RNA) provided with the spermatozoa upon fertilization or a paternal chromosome set holding a gene similar to *womanizer* may be required to start female development. Whichever gene *womanizer* codes for in *Nasonia*, its feminizing function may be provided by another gene in *Asobara* and *L. clavipes*. The regulatory complex that is responsible for the imprinting of *womanizer*-like genes need to be characterized before more conclusions can be drawn about the conservation of the upstream components in the sex determination cascades of these hymenopterans (see box 7.2 for a discussion on the possible regulatory mechanisms on epigenetic control of sex determination).