

## ***Maleness-on-the-Y (MoY) orchestrates male sex determination in major agricultural fruit fly pests***

Angela Meccariello<sup>1,\*</sup>, Marco Salvemini<sup>1,\*</sup>, Pasquale Primo<sup>1</sup>, Brantley Hall<sup>2</sup>, Panagiota Koskinioti<sup>3,9</sup>, Martina Dalíková<sup>4,5</sup>, Andrea Gravina<sup>1</sup>, Michela Anna Gucciardino<sup>1</sup>, Federica Forlenza<sup>1</sup>, Maria-Eleni Gregoriou<sup>9</sup>, Domenica Ippolito<sup>1</sup>, Simona Maria Monti<sup>6</sup>, Valeria Petrella<sup>1</sup>, Maryanna Martina Perrotta<sup>1</sup>, Stephan Schmeing<sup>7</sup>, Alessia Ruggiero<sup>6</sup>, Francesca Scolari<sup>8</sup>, Ennio Giordano<sup>1</sup>, Konstantina T. Tsoumani<sup>9</sup>, Frantisek Marec<sup>4</sup>, Nikolai Windbichler<sup>10</sup>, Javaregowda Nagaraju<sup>11</sup>, Kallare P. Arunkumar<sup>11</sup>, Kostas Bourtzis<sup>3</sup>, Kostas D. Mathiopoulos<sup>9</sup>, Jiannis Ragoussis<sup>12</sup>, Luigi Vitagliano<sup>6</sup>, Zhijian Tu<sup>2</sup>, Philippos Aris Papatianos<sup>\*\*13,14</sup>, Mark D. Robinson<sup>\*\*7</sup> and Giuseppe Saccone<sup>1\*\*</sup>

\* These authors contributed equally to this work

\*\* To whom correspondence should be addressed ([giuseppe.saccone@unina.it](mailto:giuseppe.saccone@unina.it); [mark.robinson@imls.uzh.ch](mailto:mark.robinson@imls.uzh.ch); [p.papatianos@mail.huji.ac.il](mailto:p.papatianos@mail.huji.ac.il))

1. Department of Biology, University of Naples "Federico II", 80126, Napoli, Italy
2. Department of Biochemistry, Virginia Tech, Blacksburg, United States
3. Insect Pest Control Laboratory, Joint FAO/IAEA Programme of Nuclear Techniques in Food and Agriculture, A-1400 Vienna, Austria
4. Biology Centre of the Czech Academy of Sciences, Institute of Entomology, 370 05 České Budějovice, Czech Republic
5. Faculty of Science, University of South Bohemia, 370 05 České Budějovice, Czech Republic
6. Institute of Biostructures and Bioimaging (IBB), CNR, 80134 Naples, Italy
7. Institute of Molecular Life Sciences and SIB Swiss Institute of Bioinformatics, University of Zurich, 8057 Zurich, Switzerland
8. Department of Biology and Biotechnology, University of Pavia, Pavia, 27100, Italy
9. Department of Biochemistry and Biotechnology, University of Thessaly, Larissa, Greece
10. Department of Life Sciences, Imperial College London, South Kensington Campus, SW7 2AZ, United Kingdom
11. Centre of Excellence for Genetics and Genomics of Silkmoths, Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India
12. McGill University and Genome Quebec Innovation Centre, Department of Human Genetics and Bioengineering, Montreal H3A 0G1, Canada
13. Section of Genomics and Genetics, Department of Experimental Medicine, University of Perugia, 06132, Italy
14. Department of Entomology, The Robert H Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot, 76100, Israel

## Abstract

In insects, rapidly evolving primary sex-determining signals are transduced by a conserved regulatory module producing sex-specific proteins that direct sex determination and sexual differentiation<sup>1-4</sup>. In the agricultural pest *Ceratitidis capitata* (medfly), a Y-linked maleness factor (*M*) is thought to repress the autoregulatory splicing of *transformer* (*Cetra*), which is required in XX individuals to establish and maintain female sex determination<sup>5,6</sup>. Despite previous attempts of isolating Y-linked genes using the medfly whole genome, the *M* factor has remained elusive<sup>7</sup>. Here, we report the identification of a Y-linked gene, *Maleness-on-the-Y* (*MoY*), and show that it encodes a small novel protein which is both necessary and sufficient for medfly male sex determination. Transient silencing of *MoY* in XY individuals leads to the development of fertile females while transient expression of *MoY* in XX individuals results in fertile males. Notably, a cross between these sex reverted individuals gives rise to both fertile males and females indicating that a functional *MoY* can be maternally transmitted. In contrast to the diversity of *M* factors found in dipteran species<sup>8-11</sup>, we discovered *MoY* orthologues in seven other Tephritid species spanning ~111 millions of years of evolution (Mya). We confirmed their male determining function in the olive fly (*Bactrocera oleae*) and the oriental fruit fly (*Bactrocera dorsalis*). This unexpected conservation of the primary *MoY* signal in a large number of important agricultural pests<sup>12</sup> will facilitate the development of transferable genetic control strategies in these species, for example sterile male releases or sex-ratio-distorting gene drives.

## Introduction

The Tephritidae family includes 5000 species, dozens of which are invasive and highly relevant agricultural pests of fruits and crop. *Ceratitis capitata* (medfly) is one of the most destructive members of this taxon, affecting over 200 plant species. The most successful method to control the medfly, the Sterile Insect Technique (SIT)<sup>13</sup>, involves the continuous mass-release of laboratory-reared sterilized males that suppress, or even eradicate wild populations by mating with wild females. A key determinant to the success of SIT programs, has been the close physical linkage of selectable traits to the Y-linked *M* factor in so called genetic sexing strains, that were generated to enable sex separation on a massive scale and thus the mass male release<sup>14</sup>. The development of similar strains in other Tephritid pest species however has been difficult, inhibiting the transfer of SIT or related genetic strategies. Therefore identifying the *M* factor of the medfly and documenting its evolutionary origin and conservation in related pests also holds significant promise for establishing genetic control strategies in species where efficient population control methods are currently needed.

## Results

We conducted the medfly *M* factor search by: 1) focusing on the 4-8 hours after egg laying (AEL) time window to assess embryonic transcription, the period when male sex determination is first established<sup>15</sup>; 2) producing XX-only embryonic transcriptomes used as a control in the search for Y-specific sequences; 3) developing a PacBio-based male genome assembly from the *Fam18* medfly strain<sup>6</sup>, that has a short Y chromosome; and 4) searching for conservation of putative *M* factors in another Tephritidae, the olive fly *Bactrocera oleae*. We found 19 candidate Y-chromosome derived transcripts that are expressed specifically in mixed-sex embryos, with no assigned reads from XX-only embryos (Fig. 1a; Extended Data Table 1; Supplementary Note 1). Seven of these transcripts did not map to the *Fam18* male genome assembly and were excluded. Sequence similarity searches by BLASTn showed that three out of 12 transcripts had hits to XY but not XX embryonic transcripts from the related *B. oleae*. Furthermore, one of these (DN40292\_c0\_g3\_i1) corresponds to one of the 28 transcripts previously identified by CQ approach<sup>16</sup> in a preliminary screen (Supplementary Note 2). This transcript mapped on a predicted Y-linked scaffold in the *Fam18* genome assembly (Supplementary Note 3 and 4) and lacked paralogs elsewhere in the genome,

suggesting that it may be single-copy on the Y. Functional analysis (below) confirmed that the corresponding gene is the medfly *M* factor and was thus named *Maleness-on-the-Y* (*MoY*). *MoY* is a short intronless gene located on the long arm of the Y chromosome in proximity to the centromere (Fig. 1b) in a region that contains 9 other transcriptional units, most of which are repetitive and largely overlapping on two opposite strands (Fig. 1c; Supplementary Notes 1, 3 and 4). *MoY* expression begins at 2-3h AEL, prior to embryonic cellularization, with a transient peak at 15h and becomes undetectable by 48h (Fig. 1d).

Transient silencing of *MoY* by embryonic RNAi microinjections resulted in the loss of male-specific *Cctra* transcripts in 8h AEL embryos (Extended Data Table 2; Supplementary Table 1). When we assessed *Cctra* splicing in injected 3-day-old XY larvae and adults we found a switching to the female-specific autoregulatory pattern<sup>5</sup> (Fig. 2a-b; Extended Data Fig. 1a-b; ). Among adults, 38% of the XY individuals (14/37) displayed complete phenotypic feminization and 19% (7/37) had intersex phenotypes (Fig. 2c). Phenotypic assignment of sex was based on the presence/absence of the pair of male-specific spatulae and of the female-specific ovipositor. Individuals lacking the female-specific ovipositor and male-specific spatulae, or having both, were phenotypically classified as intersexes. To evaluate the fertility of these feminized XY individuals, crosses were established to XX fertile males, which were generated by silencing of *Cctra*<sup>5</sup>. This crossing strategy simplified the tracking of any Y chromosomes that were transmitted maternally (Extended Data Fig. 3). Among the 13 recovered XY females, 2 were fertile and transmitted their Y chromosome to male progeny (Supplementary Table 2, female n. 26, Extended Data Fig. 3, Extended Data Table 2). To our knowledge, this is the first experiment in an animal species with heteromorphic XY chromosomes, in which a functional Y chromosome was transmitted maternally (Extended Data Figure 3a).

To further support the requirement of *MoY* for male sex determination, loss-of-function alleles were generated by embryo microinjections of Cas9 ribonucleoproteins<sup>16</sup> targeting a predicted 70-aa *MOY* coding sequence (Extended Data Table 2; Extended Data Fig. 4a). Indels near the sgRNA-target site of *MoY* were detected in genomes of G<sub>0</sub> XY larvae and G<sub>0</sub> XY intersexes (Extended Data Fig. 5a,b). 50% of the XY individuals were transformed into phenotypic females (2/14) or intersexes (5/14) (Extended Data Fig. 4b-c, Extended Data Fig. 5b). As above, G<sub>0</sub> XY reverted females were crossed with XX males and one of these produced female-only G<sub>1</sub> offspring as expected, including 3 XY and 18 XX females

(Supplementary Table 3; Extended Data Fig. 4d-e). Two XY G<sub>1</sub> females had frameshift-inducing deletions (10 bp and 4 bp long, respectively) in the *MoY* open reading frame (Extended Data Fig. 5a). Similarly to RNAi-induced XY females, we observed reduced fertility and bias in favour of XX offspring, possibly due to hidden sexual mosaicism of cellular clones, reduced viability of Y-carrying eggs or problems of X-to-Y pairing during female meiosis. We concluded that *MoY* is likely a protein coding gene, that is necessary for male sex determination that acts to inhibit *Cctra* female-specific splicing and/or the establishment of its autoregulation. We also concluded that the medfly Y chromosome, either with a wild type *MoY* or carrying null *MoY* mutations, can be transmitted through XY female meiosis (Extended Data Fig. 3b).

Next, we investigated whether *MoY* is sufficient for male sex determination. A 5 kb genomic fragment, encompassing the *MoY* locus and flanking regulatory regions (Fig. 1c), was injected as a linear PCR product or as circular plasmid into either XX-only or XX/XY embryos (Fig. 2d; Extended Data Table 2). As a result, male-specific *Cctra* splicing was artificially induced in XX individuals at embryonic, larval and adult stages (Fig. 2d; Extended Data Fig. 6a-c). In injection set number 4, 75% of emerging XX flies (9/12) were either fully masculinized into fertile XX males (3/12) or partially masculinized at the phenotypic or molecular levels (6/12) (Supplementary Fig. 2, Extended Data Fig. 6b). To confirm that *MoY* is a protein coding gene we performed embryo microinjections of a recombinant his-tagged MOY protein. We recovered partial phenotypic masculinization in 19% of the emerging XX adults (6/31, Extended Data Table 2, Supplementary Fig. 3). We concluded that MOY protein controls male sex determination. In contrast to M factors that have been recently discovered in other dipterans<sup>8-11</sup>, *MoY* is sufficient to fully masculinize XX individuals, making it a particularly attractive tool for translational research.

*Ceratitis* MOY showed no significant BLASTp hits to the NCBI non-redundant protein database, suggesting either novelty or a high sequence divergence. In contrast, tBLASTn searches of available assembled genomes, WGS or RNA-seq data of 14 Tephritidae species spanning 111 Mya<sup>17</sup> identified putative MOY orthologues in 8 of them (Fig. 3a, b, Supplementary Table 8). MOY exhibits 43-60% amino acid sequence similarity to the 7 MOY orthologues (Supplementary Notes 5-12). The most conserved sequence region among the 8 MOY orthologs is located in the N-terminal portion, where a well preserved hexapeptide KXNSRT occurs (Fig. 3b, MOY-R in red). We confirmed Y-linkage of *MoY*

orthologues in 4 species, namely *B. oleae*, *B. dorsalis*, *B. tryoni* and *B. jarvisi* by PCRs of male and female genomic DNA (Fig. 4a). Knock-down of *MoY* orthologs using embryonic RNAi microinjections of *B. oleae* and *B. dorsalis* embryos led to partial or full feminization of XY individuals, confirming the functional conservation of *MoY* as the male-determining Y-linked gene in other Tephritidae species (Fig. 4b; Extended Data Table 2; Extended Data Fig. 7).

## Discussion

Over the course of evolution of the insect sex determining pathway, duplication and neo-functionalization of splicing-related genes generated different types of regulators which control the splicing of *transformer*<sup>2,3,11,18,19</sup>. In contrast, our study suggests that in Tephritidae *MoY*, being unrelated to known splicing regulators or other proteins, could have emerged *de novo*<sup>20</sup>, to prevent female splicing of *transformer* and the resulting establishment of its positive feedback loop. Likely, it is the intrinsic instability of a splicing-based autoregulatory loop that makes it susceptible to interference by different mechanisms and molecules<sup>5,19,21</sup>.

In contrast to previous studies which showed that different M factors evolved in related species (*Anopheles* genus)<sup>8-10</sup> or even in the same species (*Musca domestica*)<sup>12</sup>, *MoY*'s structure and function appear to be highly conserved in Tephritidae species, which span over 111 Mya of divergence. To develop effective control strategies for agricultural pests and disease vectors, it will be important to understand the mechanisms by which *MOY*<sup>22</sup> and other short novel M proteins, such as YOB<sup>9</sup> and GUY1<sup>10</sup> in mosquitoes, orchestrate splicing of their downstream targets in the sex determining pathway.

Despite intense control campaigns, Tephritid fruit flies are diffusing in agricultural areas world wide. For example, the olive fly from the mediterranean area has recently colonized California<sup>23</sup> and the oriental fruit fly *B. dorsalis* was recently detected for the first time in Europe<sup>24</sup>. An emerging perspective from the presented *MoY* data is a remarkable plasticity of medfly development and possibly of other Tephritidae to perturbations of sex determination and differentiation. From a translational perspective, this plasticity makes *MoY* an excellent candidate for the future development of more efficient medfly genetic control strategies, such as *MoY*-based male-converting gene drives<sup>25-30</sup>. The atypical conservation of *MoY* across the Tephritid genus spanning approximately 111 Mya that we documented, makes the applied

potential of *MoY* even more appealing, because technologies based on *MoY* could be easily transferable between species in a large number of agricultural pests.

## References

- 1 Sanchez, L. Sex-determining mechanisms in insects. *Int. J. Dev. Biol.* **52**, 837-856 (2008).
- 2 Bopp, D., Saccone, G. & Beye, M. Sex determination in insects: variations on a common theme. *Sex. Dev.* **8**, 20-28 (2014).
- 3 Verhulst, E.C., van de Zande, L., Beukeboom, L.W. Insect sex determination: it all evolves around *transformer*. *Curr. Opin. Genet. Dev.* **4**, 376-83 (2010).
- 4 Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol.* **12** (2014).
- 5 Pane, A., Salvemini, M., Delli Bovi, P., Polito, C. & Saccone, G. The *transformer* gene in *Ceratitis capitata* provides a genetic basis for selecting and remembering the sexual fate. *Development* **129**, 3715-3725 (2002).
- 6 Willhoeft, U. & Franz, G. Identification of the sex-determining region of the *Ceratitis capitata* Y chromosome by deletion mapping. *Genetics* **144**, 737-745 (1996).
- 7 Papanicolaou, A. *et al.* The whole genome sequence of the Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. *Genome Biol.* **17**, 192 (2016).
- 8 Hall, A. B. *et al.* SEX DETERMINATION. A male-determining factor in the mosquito *Aedes aegypti*. *Science* **348**, 1268-1270, doi:10.1126/science.aaa2850 (2015).



- 9 Krzywinska, E., Dennison, N.J., Lycett, G.J. & Krzywinski, J. A *maleness* gene in the malaria mosquito *Anopheles gambiae*. *Science* **353**, 67-69 (2016).
- 10 Criscione, F., Qi, Y. & Tu, Z. GUY1 confers complete female lethality and is a strong candidate for a male-determining factor in *Anopheles stephensi*. *eLIFE* **5** (2016).
- 11 Sharma, A. et al. Male sex in houseflies is determined by *Mdmd*, a paralog of the generic splice factor gene CWC22. *Science* **356**, 642-645 (2017).
- 12 Qin, Y., Pains, D.R., Wang, C., Fang, Y. & Li, Z. Global establishment risk of economically important fruit fly species (Tephritidae). *PLoS One* **10** (2015).
- 13 Dyck, V.A., Hendrichs, J.P. & Robinson, A.S. *The Sterile Insect Technique: Principles and Practice in Area-Wide Integrated Pest Management* (Springer, Dordrecht, 2005).
- 14 Caceres, C. Mass rearing of temperature sensitive genetic sexing strains in the Mediterranean fruit fly (*Ceratitis capitata*). *Genetica* **116**, 107-116 (2002).
- 15 Hall AB, Qi Y, Timoshevskiy V, Sharakhova MV, Sharakhov IV, Tu Z. Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics*. 2013 Apr 23;14:273.
- 16 Meccariello, A. et al. Highly efficient DNA-free gene disruption in the agricultural pest *Ceratitis capitata* by CRISPR-Cas9 ribonucleoprotein complexes. *Sci. Rep.* **7**, 10061 (2017).
- 17 Krosch, M.N. et al. A molecular phylogeny for the Tribe Dacini (Diptera: Tephritidae): systematic and biogeographic implications. *Mol. Phylogenet. Evol.* **64**, 513-523 (2012).
- 18 Cline, T.W. et al. Evolution of the *Drosophila* feminizing switch gene *Sex-lethal*. *Genetics* **186**, 1321-1336 (2010).



- 19 Bopp, D. About females and males: continuity and discontinuity in flies. *J. Genet.* **89**, 315-23 (2010).
- 20 Schlötterer, C. Genes from scratch - the evolutionary fate of *de novo* genes. *Trends Genet.* **4**, 215-219 (2015).
- 21 Saccone, G., Salvemini, M. & Polito, L.C. The *transformer* gene of *Ceratitis capitata*: a paradigm for a conserved epigenetic master regulator of sex determination in insects. *Genetica* **139**, 99-111 (2011).
- 22 Saccone, G. et al. Male-specific phosphorylated SR proteins in adult flies of the Mediterranean fruitfly *Ceratitis capitata*. *BMC Genet* **15 Suppl. 2**, S6 (2014).
- 23 Daane, K.M. & Johnson, M.W. Olive fruit fly: managing an ancient pest in modern times. *Annu. Rev. Entomol.* **55**, 151-169 (2010).
- 24 Nugnes, F., Russo, E., Viggiani, G. & Bernardo, U. First Record of an Invasive Fruit Fly Belonging to *Bactrocera dorsalis* Complex (Diptera: Tephritidae) in Europe. *Insects* **9** (2018).
- 25 Gantz VM, Bier E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science*. 348(6233):442-4 (2015). doi: 10.1126/science.aaa5945
- 26 Kyrou, K. et al. A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. *Nat. Biotechnol.* **36**, 1062-1066 (2018).
- 27 KaramiNejadRanjbar, M. et al. Consequences of resistance evolution in a Cas9-based sex conversion-suppression gene drive for insect pest management. *Proc. Natl. Acad. Sci. USA* **115**, 6189-6194 (2018).
- 28 Kandul, N.P. et al. Transforming insect population control with precision guided sterile males with demonstration in flies. *Nat. Commun.* **10**, 84 (2019).
- 29 Krzywinska, E. & Krzywinski, J. Effects of stable ectopic expression of the primary sex determination gene *Yob* in the mosquito *Anopheles*

*gambiae*. *Parasit. Vectors* **11**, 648, doi:10.1186/s13071-018-3211-z (2018).

- 30 Papathanos, P. A., Windbichler, N. Redkmer: An assembly-free pipeline for the identification of abundant and specific X-chromosome target sequences for X-shredding by CRISPR endonucleases. *CRISPR J.* **1**, 88-98 (2018).

**Supplementary Information** is available in the online version of the paper.

### Acknowledgements

This paper is dedicated to the memory of JN who passed away during the first phase of the project. We thank Carlos Caceres and Sohel Ahmad (FAO-IAEA, Seibersdorf, Austria), for providing *B. oleae* strain, material and methods to rear them. We thanks Bernardo Carvalho (University of Rio de Janeiro) for suggestions on bioinformatics, Jennifer Morrow (Western Sydney University) for providing *B. jarvisi* and *B. tryoni* fly samples, and Daniel Bopp (University of Zurich) for discussions, suggestions, and for a generous critical reading and editing of the manuscript We thank Serena Aceto (University of Naples Federico II) for suggestions on the phylogenetic analysis and protein alignments.

RNA and DNA Sequencing costs were covered by MDR (Swiss National Science Foundation (project numbers 31003A\_143883, 310030\_175841 and support from the University of Zurich, Research Priority Program Evolution in Action for SS). This research was supported by “Ricerca Dipartimentale” grants to GS and MS (Department of Biology, Federico II University of Naples, Italy, 2015-2018), by the grant STAR2013\_25 to MS from University of Naples Federico II and Compagnia di San Paolo, Naples, ITALY, in the frame of Programme STAR2013 (Sostegno Territoriale alle Attività di Ricerca) and by PhD fellowships to AM and PP from PhD program in Biology (Department of Biology, University of Naples Federico II, Italy). NW was funded by the BBSRC under the research grant BB/P000843/1. ZT was supported by NIH grants AI121284 and AI123338. PAP was supported by a Rita Levi Montalcini award from the Ministry Education, University and Research (MIUR—D.M. no. 79 04.02.2014). JR was supported by Compute Canada

Resource Allocation Project (WST-164-AB), and Genome Canada Genome Innovation Node (244819) grants. We would like to thank Haig Djambadzian and Anthony Bayega (McGill) for providing data and manuscript edits. This study was benefited from discussion during the research coordination meetings of several Coordination Research Projects (CRP) supported by the International Atomic Energy and particularly of the CRP on "Comparing Rearing Efficiency and Competitiveness of Sterile Male Strains Produced by Genetic, Transgenic or Symbiont-based Technologies".

### Author Contributions

GS and JN initiated the project. GS conceived the production of XX-only embryos for mRNA-seq, high-throughput functional tests by RT-PCR (after injection of dsRNA or DNA) in early embryos to identify the M factor, by embryos injections of recombinant MOY protein, the use of FAM18 strain carrying a shorter Y, and performed the identification of *MoY* Tephritidae orthologues. MS improved the crossing scheme to produce XX-only embryos. GS, MDR, MS, ZT, BH, AK, JN, PAP, NW designed and performed bioinformatics analyses, with contributions from AM and PP. ZT, BH performed preliminary CQ analysis. MS, VP and AM prepared RNA and DNA for sequencing. MDR performed mRNA sequencing, DE and CQ analyses, PacBio FAM18 sequencing, and Canu assembly together with SS. MS performed the medfly 4-8h embryonic transcriptome assembly and the Tephritidae species transcriptome assemblies. MS developed the local web-tool with graphical interface utilized for all the BLAST searches of this paper. AM and GS selected *MoY* as a first M candidate from the list of putative male-specific transcripts by MDR and the list of novel putative male-specific transcripts by ZT and BH (Supplementary Note 2). AM performed eRNAi, CRISPR/Cas9, and *MoY* DNA injections and RT-PCR analyses demonstrating *MoY* function. PP, AG, MAG, FF, DI and MMP maintained the strains, performed crosses and DNA/RNA molecular analyses. LV and AR purified recombinant MOY protein and performed structural/similarity analyses to protein databases. PP performed MOY protein embryos injections. FM and MD performed *in situ* hybridization of *MoY*. KM, JR, KTT, MEG performed qRT-PCR analysis of *MoY* in *C. capitata* and *MoY* expression analysis in *Bactrocera oleae*. KM and JR provided *Bactrocera oleae* genome assembly data. AM, FS, PK and KB performed *MoY* RNAi on *MoY* orthologues of *Bactrocera oleae* and *Bactrocera dorsalis*. PK and KB performed molecular analysis on reverted XY females in

these two species. GS, AM, MS and PAP prepared the figures. All of the authors discussed the data. PAP, SMM, EG and LV provided essential reagents. GS wrote the manuscript with intellectual input from all authors, especially PAP, MDR, MS, and LV. AM and MS contributed equally to the work. GS supervised the project.

**Author Information** The 12 Kb *MoY* sequence region has been deposited in GenBank under accession number MK330842. Reprints and permissions information is available at. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to GS ([giuseppe.saccone@unina.it](mailto:giuseppe.saccone@unina.it)), MDR, ([mark.robinson@imls.uzh.ch](mailto:mark.robinson@imls.uzh.ch)), or PAP ([p.papathanos@mail.huji.ac.il](mailto:p.papathanos@mail.huji.ac.il)).

## Methods.

### Insects rearing and strains

The following medfly strains were used in this study: 1) the wild-type strain *Benakeion* developed by P. A. Mourikis (Benakeion Institute of Phytopathology, Athens, Greece); 2) the Fam18 strain<sup>6</sup> (see below); 3) the *Vienna8* strain<sup>12</sup>, carrying a Y chromosome marked with translocated wild type allele of a *white pupae* mutation. The *C. capitata Benakeion* strain was reared in standard laboratory conditions at 25 °C, 70% relative humidity and 12:12 h light–dark regimen. Adult flies were fed yeast/sucrose powder (1:2). Eggs were collected in trays filled with distilled water and transferred to larval food (Piccioni lab, Italy) after hatching. Pupae were collected and stored in dishes until eclosion.

The *Bactrocera dorsalis* Saraburi strain originated from Thailand and was maintained at the FAO/IAEA Insect Pest Control Laboratory (IPCL, Seibersdorf, Austria) for 84 generations before use in these experiments. Flies were reared at 25 ± 1 °C, 60 ± 5% RH and a photoperiod of 12h light : 12h dark, and fed with a standard laboratory adult diet containing sugar and hydrolyzed yeast in a 3:1 ratio and water *ad libitum*. Eggs were collected and transferred to larval medium containing 28% wheat bran, 7% brewer's yeast, 13% sugar, 0.21% sodium benzoate, 0.21% nipagin, 1% HCl, 50% water.

The *Bactrocera oleae* Greece-Lab strain originated from a stock in the Department of Biology, 'Demokritos' Nuclear Research Centre, Athens, Greece, and was maintained at the FAO/IAEA Insect Pest Control Laboratory IPCL for 177 generations. Flies were reared at 25

$\pm 1$  °C,  $60 \pm 5\%$  RH, a photoperiod of 14h light : 10h dark, and fed with a standard laboratory adult diet consisting of 75% sugar, 19% hydrolyzed yeast and 6% egg yolk powder. Eggs were collected and transferred to larval diet containing 550 mL water, 20 mL Extra Virgin Olive Oil, 7,5 mL Tween® 80 emulsifier, 0.5 gr potassium sorbate, 2 gr nipagin, 20 gr sugar, 75 gr brewer's yeast, 30 gr soy hydrolysate, 4.5 mL HCl 36% and 275 gr cellulose powder.

### **Production of XX-only embryos.**

Embryonic RNAi targeting *Cctra* was used to produce male-only progeny including XX fertile reverted males, as previously described<sup>5</sup>. 40 males were individually crossed 3 wild type virgin females each. 4-8h AEL embryos were collected in 40 separated samples for RNA extractions and RT-PCR assays using Y-specific marker (*Cclap*). 22 RNA samples corresponded to XX-only progeny from individual crosses with XX fertile males and were pooled in 3 sets for sequencing.

### **RNA sequencing, *de novo* transcriptome assembly and differential gene expression**

Total RNA was prepared from 4-8 h AEL (when male sex determination occurs<sup>31</sup>) *C. capitata Benakeion* strain embryos of mixed sexes (XX/XY) and of female-only sex (XX; see above), in three biological replicates each, with TRIzol reagent (Thermo Fisher Scientific) according to the manufacturer's protocol. The integrity and purity of extracted total RNA were assessed using the NanoDrop 2000c (Thermo Fisher) and the Agilent 2100 Bioanalyzer with RNA 6000 Nano kit (Agilent, Santa Clara, CA, US). All RNA samples had a A260/280 ratio higher than 2.1 and a RIN value higher than 7.8. Total RNAs were depleted of ribosomal RNA with Ribo-Zero rRNA Removal kit (Epicentre). Libraries for sequencing, with an average insert size of 300 bp, were generated with the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). Illumina stranded and paired-end (PE) sequencing was performed at NXT-GNT (Belgium) yielding a total of 85,832,487 XY/XX and 94,085,461 XX-only PE 100 bp-long reads (BioProject PRJNA434819 in SRA archive). A transcript catalogue was *de novo* produced using the whole 4-8 hrs Illumina read data set, concatenated into two paired FASTQ files and the Trinity assembler<sup>32,33</sup>, run with default parameters and `-SS_lib_type RF -jaccard_clip -normalize_reads` flags set, after adapter removal and quality trimming using Trimmomatic v.0.33<sup>34</sup> and mitochondrial and ribosomal reads depletion, as described<sup>35</sup>. Assembly statistics are presented in Supplementary Table 4. Transcript-level quantifications for each sample were done against this catalogue

using Kallisto<sup>36</sup> (Supplementary Table 5). Differential gene expression was performed using edgeR<sup>37</sup>, with cut-off values of FDR < 0.05 and logFC > 0 (Supplementary Table 6).

### ***Fam18* strain description**

Due to the lower complexity of its shorter Y chromosome, the *Fam18* strain (kindly provided by Dr Franz, FAO/IAEA, Austria), was used for PacBio sequencing of male genomic DNA. This strain originated from a study by Willoefth and Franz<sup>6</sup>, who used cytogenetic approaches, including *in situ* hybridization with Y-specific repetitive probes<sup>38</sup> and Y-autosome translocations, to map the male-determining factor region on the long arm of the *C. capitata* Y chromosome, toward the centromere. This region represents approximately 15% of the entire medfly Y chromosome. In *Fam18*, males carry an internal deletion of the long arm of the Y chromosome but not Y-autosomal translocations, and this Y chromosome lacks a segment between the M region and a point toward the tip (Franz, G., FAO-IAEA Pest Control Unit, person. comm. to GS, 2010). To verify that in the *Fam18* males there is a reduced complexity of the Y chromosome due to the large deletion of its long arm, we mapped, using the Bowtie aligner<sup>39</sup>, the *Fam18* males and females Illumina reads against four Y-specific repetitive elements, known to be widely distributed all over the medfly Y chromosome<sup>6,38</sup>. *Fam18* read counts are significantly lower for three out of four Y-linked repetitive elements compared to the medfly ISPRA standard strain read counts (Supplementary Table 7). Conversely, the read counts for the control gene are comparable between the two strains.

### **Illumina short read WGS of *C. capitata Fam18* strain**

Five adult males and five adult females from the *C. capitata Fam18* strain were collected and genomic DNA was extracted from the two sexed pools separately, as described below<sup>40</sup>. Each pool was sequenced with 1 lane of Illumina HiSeq2500 resulting in 243,152,822 male and 248,886,022 female 2x126-bp PE reads, respectively (BioProject PRJNA435549 in SRA archive).

### **Illumina short read WGS, trimming, error-correction and assembly**

The Illumina raw short-read data for the males had approximately 126X coverage (assuming a genome size of 485Mbp). Trimmomatic v0.35<sup>34</sup> removed the Illumina adapters in paired-end mode, trimmed by quality values in maxInfo mode with target\_length=0 and

strictness=0.4 parameters, and selected for a remaining read length of at least 50 bases. The resulting trimmed reads were used for all purposes except for the error correction of the PacBio reads. The trimmed reads had a coverage of 62X for the first and 59X for the second read.

For the PacBio correction, we first corrected the trimmed short reads with BFC<sup>41</sup> (r181) using k-mer length=33 and genome size=485Mb as parameters. Meraculous<sup>42</sup> assembled those corrected short reads into unitigs with genome size=485Mb, num\_prefix\_blocks=8, and automatic k-mer length identification. To determine the library parameters for the corrected short reads, we mapped the reads to the Ccap01172013 genome from Baylor College<sup>7</sup> using bowtie2<sup>43</sup> v2.2.5 with a maximum insert length of 1000. This resulted in a mean insert size of 364 bp with a Gaussian standard deviation of 79 bp. The average read length of the corrected reads was 121 bp. The corrected reads and the unitigs were then used to correct the PacBio reads.

### **PacBio Whole Genome Sequencing (WGS) of *C. capitata* Fam18 strain**

Five adult males from the *C. capitata* Fam18 strain were collected and genomic DNA was extracted using the Holmes-Bonner protocol<sup>40</sup>. Extracted genomic DNA was resuspended in 10 mM Tris-HCl, pH 8.5, and its integrity and purity were assessed using the NanoDrop 2000c (Thermo Fisher Scientific) and 1% agarose gel electrophoresis, resulting in A260/280 ratio of 1.8. Genomic DNA was sheared by g-TUBE (Covaris) and a BluePippin instrument (Sage Science) was used to select fragments from 15-20 Kb in length. A large insert library was prepared using Pacific Biosciences recommended protocols and P6-C4 reagents. The resulting sample was sequenced at the Functional Genomics Centre Zurich (Switzerland) with 20 SMRT cells on a Pacific Biosciences RSII instrument. The PacBio SMRT cell reads have been submitted to the NCBI SRA archive with the following accession number: BioProject PRJNA435534.

### **PacBio WGS error-correction and assembly**

PacBio DNA sequencing using P6-C4 chemistry for 20 SMRT cells was conducted; extraction of sequences into FASTQ files was done by Dextractor (<https://github.com/thegenemyers/DEXTRACTOR>) with a minimum sequence quality parameter set to 400 produced ~32X genome coverage (assuming a genome size of 485Mbp) with an average read length of 5,526 bp and a N50 length of 8,656 bp, and a maximum length



of 68,782 bp. We combined all SMRT cells and then split the data into chunks of 400 Mb for cluster processing. The error correction was done by Proovread<sup>44</sup> using the corrected first reads from the PE short read data and the unitigs. We did not use the second reads as they were of lower quality and the additional coverage would have slowed down the error correction without improving the result. The only additional parameter set was coverage=62. This resulted in a total of ~19X corrected sequences with an average read length of 2,855 bp, N50 length of 4,695 bp, and maximum length of 33,053 bp. This set of corrected reads was used to build a *de novo* *Fam18* genome assembly using Canu v1.0 with -pacbio-corrected set and genomeSize=0.45g. This resulted in 12,165 assembled contigs with a total length of 496 Mb, an average read length of 40,769 bp, N50 length of 62,964 bp, and maximum length of 1.0 Mbp.

### **Bioinformatics for *MoY* gene structure**

BLASTn search of the PacBio Canu assembly, using medfly *C. capitata MoY* 681 nt-long transcript sequence (including 5' and 3' UTRs) as DNA query (Genbank acc. num. MK165756), identified a highly similar corresponding genomic sequence (95% identity). The *MoY* putative coding region of the transcript is 99% identical in the PacBio genomic sequence, with only 2 SNPs, with the second inducing a conservative amino acid substitution at position 63 (I->M) (Genbank acc. num. MK165755).

### **Probe preparation and labelling**

1.5 kb fragment of *MoY* plasmid was amplified by PCR using the primers *New\_MoY\_F1/New\_MoY\_R*. The reaction with a final volume of 20 µl contained 22 ng of plasmid DNA, 0.2 mM of each dNTP, 1 µM of each primer, 1× Ex *Taq* buffer, and 1U Ex *Taq* DNA Polymerase (TaKaRa, Otsu, Japan). The PCR product was purified using Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA) according to manufacturer instructions.

Probe labelling was performed using an improved nick translation procedure<sup>45</sup> with some modifications. The modified 40 µL reaction contained 500 ng of unlabeled DNA, 50 µM of each dATP, dCTP, and dGTP, 10 µM of dTTP, 20 µM digoxigenin-11-UTP (Roche Diagnostics, Mannheim, Germany), 1× nick translation buffer (50 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 0.005% BSA), 10 mM β-mercaptoethanol, 1.25×10<sup>-4</sup> U/µl DNase I and 1 U/µl DNA polymerase I (both ThermoFisher Scientific, Waltham, MA, USA). The reaction was

incubated at 15 °C for 40 min and then purified using an illustra Sephadex G-50 column (GE Healthcare, Buckinghamshire, UK).

### **Chromosome preparation**

Spread preparations of mitotic chromosomes were made from the brain (cerebral ganglia) of third instar larvae following a previously described method<sup>46</sup> with slight modifications. Briefly, the brain was dissected in a physiological solution, transferred into hypotonic solution (0.075 M KCl) for 10 min, and then fixed for 15 min in freshly prepared Carnoy's fixative (ethanol:chloroform:acetic acid, 6:3:1). The fixed tissue was spread in a drop of 60% acetic acid on the slide at 45 °C using a heating plate. Then the preparations were passed through a graded ethanol series (70%, 80%, and 100%, 30 s each), air dried, and stored at -20 °C. Before further use, the preparations were again dehydrated in the ethanol series, immediately after removal from the freezer.

### **Fluorescence *in situ* hybridization with tyramide signal amplification (TSA-FISH)**

The TSA-FISH was performed according to published protocols<sup>46,47</sup> with some modifications. Slides were pre-treated with 200 µg/mL RNase A in 2× SSC at 37 °C for 1 h, with 0.01 M HCl at 37 °C for 20 min, 1% H<sub>2</sub>O<sub>2</sub> in 2× SSC at room temperature (RT) for 30 min, and 5× Denhardt's solution (0.1% Ficoll 400, 0.1% polyvinylpyrrolidone, and 0.1% bovine serum albumin in ultrapure water) at 37 °C for 30 min. The probe cocktail containing 50% deionized formamide, 10% dextran sulphate, and 0.5 ng/µL of the probe in 2× SSC was applied to chromosome preparation, covered with cover slip and denatured at 70 °C for 8 min. Hybridization was carried out at 37 °C overnight.

After hybridization, the slides were washed three times in 50% formamide in 2× SSC at 46 °C for 5 min, three times in 2× SSC at 46 °C for 5 min, three times in 0.1× SSC at 62 °C for 5 min, and in TNT buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.05% Tween-20) at RT for 5 min. The slides were blocked in TNB buffer (0.1 M Tris-HCl pH 7.5, 0.15 M NaCl, 0.5% Blocking Reagent; PerkinElmer, Waltham, MA, USA) at RT for 45 min. The probe was detected by anti-digoxigenin-POD (Roche Diagnostics) diluted 1:20 in TNB buffer.

Tyramide amplification was carried out using the TSA<sup>TM</sup> Plus Fluorescein System (PerkinElmer) according to the manufacturer's instructions. The slides were incubated with tyramide working solution at RT for 20 min and then washed as described previously<sup>46</sup>. Chromosomes were counterstained with 0.5 µg/mL DAPI (4',6-diamidino-2-phenylindole; Sigma-Aldrich, St. Louis, MO, USA) in 1× PBS at RT for 15 min. Finally, the slides were mounted in antifade based on DABCO (1,4-diazabicyclo(2.2.2)-octane; Sigma-Aldrich). Chromosome preparations were observed in a Zeiss Axioplan 2 microscope (Carl Zeiss, Jena, Germany). Black-and-white images were captured separately for each fluorescent dye with a monochrome CCD camera XM10 using cellSens Standard software version 1.9 (both Olympus Europa Holding, Hamburg, Germany). The images were pseudo-coloured and merged using Adobe Photoshop CS5 (Adobe Systems, San Jose, CA, USA).

### **Evaluation of *MoY* expression via qRT-PCR: RNA isolation, cDNA synthesis and primer design**

For the validation of the *MoY* gene expression, RNA was extracted from several developmental stages and tissues of the medfly. A pool of ~50 eggs were removed from the incubator at different time intervals throughout embryonic development. Each embryonic sample refers to the time after oviposition, at which eggs were collected in Extrazol (Nanogen Advanced Diagnostics, Turin, Italy). In total, 10 different embryonic time points, i.e. 1h, 2h, 3h, 4h, 5h, 6h, 7h, 8h, 15h and 48h, were examined in three biological replicates. Regarding the other developmental stages, a pool of 10 per sample was collected for 1<sup>st</sup> and 3<sup>rd</sup> instar larvae respectively, whereas a pool of 5 per sample was collected for pupae, female adults, male adults, as well as for testes. These aforementioned samples were examined in either three or two biological replicates (in the case of male adult tissues).

Following RNA extraction, genomic DNA was removed using a DNase treatment with 1.0 unit of TURBO<sup>TM</sup> DNase (Invitrogen, CA) according to manufacturer's instructions. The total amount of DNA-free RNA obtained from each tissue was reverse transcribed using the PrimeScript RT reagent Kit (TaKaRa, Japan) according to manufacturer's instructions. Reverse transcription was conducted at 37°C for 15 min and 85°C for 5 min. The resulting cDNA was used in the subsequent qPCR reactions. Specific primers were designed by IDT primer-designing software PrimerQuest on the Integrated DNA Technologies (IDT) website (<http://www.idtdna.com>) for the amplification of *CcMoY* (**Supplementary Methods** Table 1). A quantitative Real-time PCR (qRT-PCR) approach was performed to analyze changes in

*MoY* transcription. The expression levels were obtained relatively to the suitable housekeeping gene based on the evaluating reference genes list of each species<sup>48</sup>. The *β-tub* gene was used to normalize expression in eggs, whereas *Rpl19* for larvae, pupae, male and female adults respectively (Supplementary Methods Table 2). The qRT-PCR cycling conditions were: polymerase activation and DNA denaturation step at 95 °C for 3 min, followed by 50 cycles of 10 s at 95°C, 20 s at the annealing temperature, and 30 s at 72°C. Each qPCR run was followed by a melting curve analysis (gradual increase of temperature between 55 and 95°C) to examine the specificity of each reaction. For all primers, a single gene specific peak was observed, confirming the absence of non-specific products or primer dimers. Each reaction contained 5µl from a dilution 1:10 of the cDNA template, 1X KAPA SYBR FAST Universal qPCR Master Mix (Kapa Biosystems) and 300nM of each primer, to a final volume of 15µl. The reactions were carried out on Bio-Rad Real-Time thermal cycler CFX96 (Bio-Rad, Hercules, CA, USA) and data analysis was performed using the CFX Manager™ software.

### **Embryonic RNA interference**

*CcMoY*, *BdMoY*, and *BoMoY* fragments were PCR amplified using genomic DNA from a male adult of *C. capitata*, *B. dorsalis* and *B. oleae* (List of primers in Supplementary Methods Table 1), introducing a T7 promoter sequence at each end. PCR was performed with Phusion® High-Fidelity DNA Polymerase (New England Biolabs) according to manufacturer instructions. PCR product was purified with a phenol/chloroform purification and eluted in ddH<sub>2</sub>O for a final concentration of 1 µg/µL. *In vitro* transcription of *MoY* dsRNA for all the species was obtained using Ambion MEGAscript® RNAi kit T7 RNA polymerase, following manufacturer instructions. Final product was checked on 1% agarose gel electrophoresis and quantified on NanoDrop 2000c (Thermo Fisher Scientific). The dsRNA was eluted from the purification column using elution buffer, containing 5 mM KCl and 0.1 mM PO<sub>4</sub>, pre-heated at 98 °C, to get a final dsRNA yield of 1.5 µg/µL.

Embryos were collected 45 min AEL, placed on double-stick tape, hand-dechorionated, dehydrated in a Petri dish containing calcium chloride, covered with Halocarbon oil 27 (Sigma, H8773) for the three species. A solution of 1 µg/µL dsRNA was loaded inside the needle using a P10 with Eppendorf Microloader® capillary tips. For the microinjections in *Ceratitidis capitata*, an inverted contrast microscope Leica DM IRB® was used. The needles

were obtained from Boron-Silicate cylinders with a diameter of 0.5 mm pulled with a puller machine (Narishige). The injections were conducted in the posterior pole of the embryos, with the needle entering the embryo for about 30-50% of its length, and the pressure was generated using an oil pump. Injections in *Bactrocera spp.* were done into the posterior end of the eggs using XenoWorks® Digital Microinjector and Micromanipulator System (Sutter Instrument) and Quartz needles (#QF100-70-10), pulled in P-2000 Laser-Based Micropipette Puller (Sutter Instrument).

### **RNA extractions and RT-PCR on *C. capitata***

Total RNA was extracted from pools of embryos, larvae, and intersexes, and male and female adults using TRIzol Reagent (Thermo Fisher Scientific) following manufacturer instructions. Oligo-dT-primed cDNA was prepared from DNase I-treated total RNA using EuroScript® m-MLV reverse transcriptase (Euroclone®). RT-PCR expression analysis was performed using primers described in Supplementary Methods Table 1.

### **Splicing of *Bdtra* and *Botra***

Total mRNA was extracted from eggs 12 hours after injection and 3rd instar injected larvae using TRIzol Reagent (Thermo Fisher Scientific) and treated with TURBO™ DNase (Invitrogen) according to manufacturer's instructions. cDNA was synthesized using SuperScript™ III First-Strand Synthesis System (18080051) following manufacturer's instructions and PCR amplification of *tra* was performed using Qiagen Taq PCR Master Mix Kit (201443) according to manufacturer's instructions. RT-PCR expression analysis was performed using the primers reported in Supplementary Methods Table 1.

### **Molecular karyotyping of *Bactrocera* adults**

DNA was extracted using EXTRACTME DNA TISSUE kit (EM03) following manufacturer's instructions. PCR was performed using Qiagen Taq PCR Master Mix Kit (201443) according to manufacturer instructions. The primers for the amplification of the Y-specific fragment were *BdMoY-F* and *BdMoY-R* for *B. dorsalis*; *BoMoY-F* and *BoMoY-R* for *B. oleae*. PCR analysis was performed using primers described in Supplementary Methods Table 1.

### **CRISPR and RNP complex assembly in *C. capitata***

sgRNA was designed using CHOPCHOP<sup>49</sup>. Following, production of template for sgRNA was performed, with minor modifications, as described<sup>16</sup> with CRISPR-*MoY-F* and the invariant reverse primer (PAGE-purified, Life Technologies) (Supplementary Methods Table 1). sgRNA was synthesized according to instructions of the Megascript<sup>®</sup> T7 kit (Ambion) with 1 µg of template and a 5' flanking T7 promoter as starting material. After RNA synthesis, template was removed by incubating with TurboDNase<sup>®</sup> (Ambion) for 15 min at 37 °C. Cas9 was expressed as a His-tagged protein and purified from bacteria as described<sup>16</sup>. Prior to the injection, the RNP complex was prepared by mixing 1.8 µg of purified Cas9 protein with approximately 200 ng of sgRNA in a 5 µL volume containing 300 mM KCl<sup>16</sup>. The mix was incubated for 10 min at 37 °C. A glass needle was filled with the pre-loaded sgRNA-Cas9 mix and the injection was performed into the posterior end of embryos collected 45 min AEL as described for RNA interference in *C. capitata*. 48 hours after the injections, the slides are inspected for hatched larvae under a stereomicroscope. Hatched larvae are transferred to Petri dishes containing larval food.

### **Amplification and cloning of the 5 Kb fragment of *CcMoY***

PCR on the 5 Kb fragment of *CcMoY* was carried out with the LongAmp<sup>®</sup> Taq DNA Polymerase following manufacturer's instructions in a final volume of 100 µL using the primers *Gen\_CcMoY\_F* and *Gen\_CcMoY\_R* (Supplementary Methods Table 1). The 5 kb *CcMoY* fragment was then cloned in the pGEM<sup>®</sup>-T Easy vector (Promega) following manufacturer's instructions. The PCR product and the plasmid with the 5 Kb insert were purified with phenol:chloroform (1:1), precipitated in ethanol and eluted with injection buffer (containing 5 mM KCl and 0.1 mM PO<sub>4</sub>) for a final concentration of 1 µg/µL to be injected in *C. capitata* embryos.

### **MOY protein expression, purification and characterization**

The *MoY* gene was amplified using a male adult *C. capitata* genomic DNA as template with two synthetic primers carrying the *NcoI* and *XhoI* restriction sites: *CcOrf2fw (NcoI)/CcOrf2rv (XhoI)*. The fragment was then cloned into the pETM-13 expression vector. The resulting gene contained an extra C-terminal tag of six histidine residues. The MOY protein was then heterologously expressed in *Escherichia coli* BL21 (DE3) cells. MOY was purified from inclusion bodies after removing the soluble proteins by sonicating the bacterial resuspension in a native buffer. The pellet was then resuspended in a denaturing buffer



containing 8 M Urea, 50 mM Tris-HCl and 200 mM NaCl (pH 8.0) and left at room temperature for 2 hours to facilitate its solubilization. After the centrifugation, the supernatant was loaded on a Ni-NTA resin (Qiagen) equilibrated with the denaturing buffer. The protein was washed with ten volumes of this buffer and then was eluted with a solution containing a high concentration of imidazole (150-300 mM pH 8.0). The homogeneity of the eluted protein was evaluated by SDS-PAGE analysis. MOY protein was refolded in solution containing 20 mM ethanolamine and 0.5 M L-arginine (pH 9.0-10.0) using rapid dilution (1:50 v/v) with three additions of the same volume over a 24h period. The protein was concentrated using Amicon centrifugal devices (Millipore) up to a concentration of 70 µg/mL. The molar mass of the refolded MOY was determined by mass spectrometer (ESI-TOF) (9250.28 Da) in agreement with that computed from the tagged sequence (9251.7 Da).

### ***Bactrocera* spp. transcriptome assemblies and *MoY* orthologs search**

Illumina RNA-seq data for 14 *Bactrocera* species were downloaded from NCBI SRA archive and *de novo* assemblies were produced using Trinity assembler<sup>32</sup> with default parameters and -normalize\_reads and -jaccard\_clip flags set. Assembly statistics and accession numbers of the utilized data set are listed in Supplementary Table 8.

tBLASTn search of the *B. oleae* transcriptome, using MOY as protein query, identified a transcript encoding a putative protein with 36% identity over a 64 aa-long region (*BoMoY*, 70 aa; MK165746) (Supplementary Notes 5-12). TBLASTN search, using BoMOY on *B. oleae* NCBI WGS database, failed to find the full-length corresponding *BoMoY* gene, but revealed the presence of 4 genomic sequences (Sequence IDs: JXPT01043932.1; LGAM01008500.1; LGAM01009849.1; LGAM01000393.1) encoding for truncated 15-41 aa-long *BoMoY* paralogous sequences and showing 37%-100% aa identity to BoMOY protein. The identified *B. oleae* *MoY* ortholog was utilized to search by tBLASTn for *MoY* orthologues in the other thirteen assembled *Bactrocera* transcriptomes. *MoY* orthologs were found only in *B. jarvisi* (MK165748) and *B. zonata* (MK165752). As we reasoned that *MoY* transcripts could be underrepresented in assembled transcriptomes for both technical limitations and low gene expression levels, we searched for *MoY* orthologs by tBLASTn searches at NCBI on SRA and WGS databases. Sequences encoding full-length *MoY* orthologs were found in both *B. tryoni* (JHQJ01009763.1) and *B. latifrons*, showing 63% aa identity with *BoMoY* (MIMC01001452.1). tBLASTn analyses revealed also the presence in *B. tryoni* of a shorter *MoY* paralogue (50 aa long) in the same genomic contig, 4 Kb apart from the first *BtMoY*



copy (70 aa long). The two *BtMoY* paralogues showed 90% protein identity over a 50 aa-long region. Similarly, in *B. latifrons* we have found a truncated paralogue *BlMoY* gene in a different genomic contig (MIMC01001198.1), encoding for truncated 36 aa-long *BlMoY* sequence, with 78% identity to *BlMoY* and 63% to *BoMoY*. Finally, by tBLASTn search we identified a SRA read from *B. dorsalis* transcriptome analysis data (SRR316210) encoding part of the putative *BdMoY*, corresponding to the amino-terminus. We performed PCR on cDNA from 5-8 hours old embryonic RNA using primers *Bd\_MoY+*/*Bd\_MoY-* designed on the corresponding paired Illumina reads (Supplementary Note 8), leading to the isolation of the *BdMoY* full-length coding region.

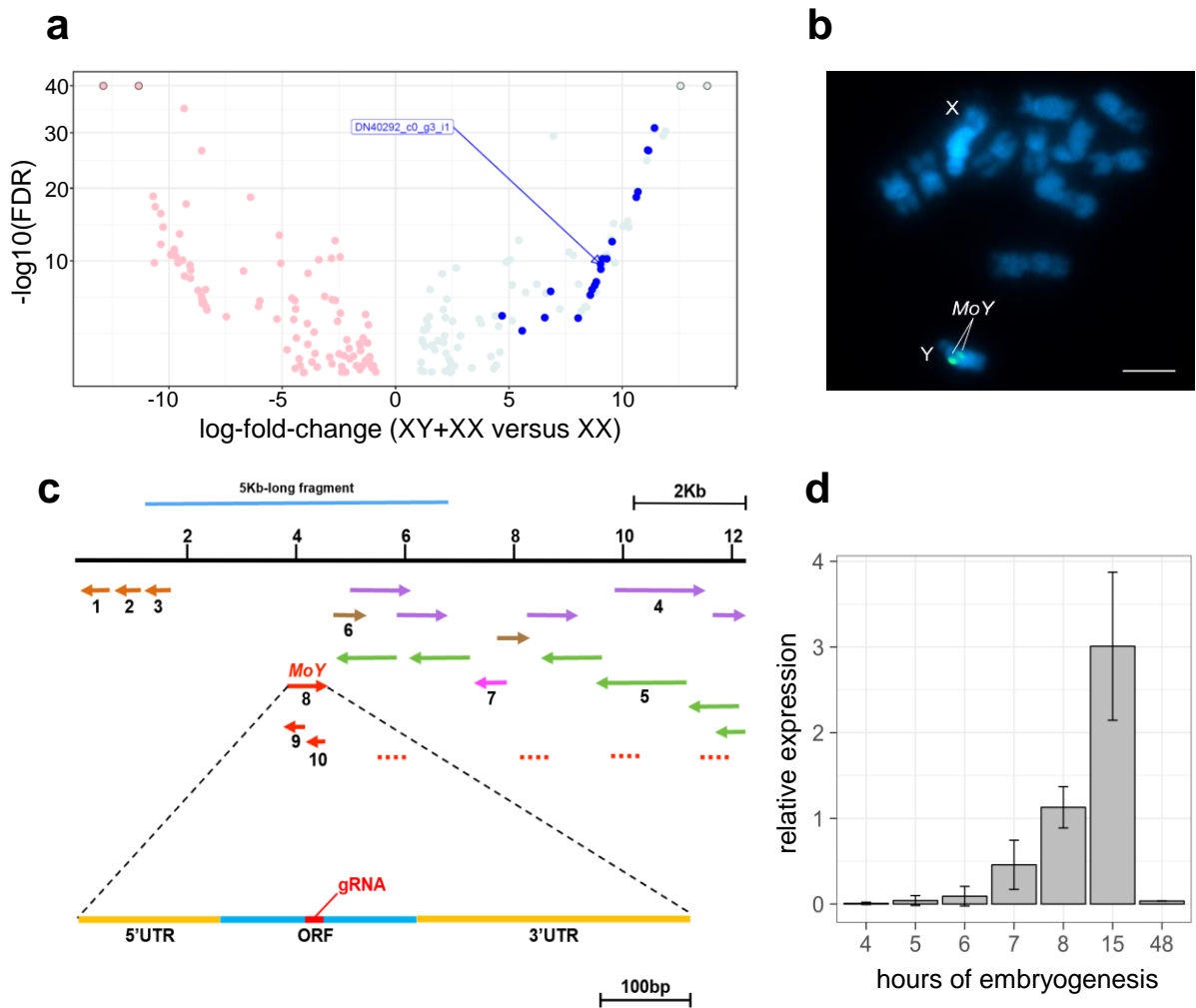
We validated the *MoY* male-specific Y-linked genomic localization by amplification on genomic DNA of *B. oleae*, *B. dorsalis*, *B. tryoni* and *B. jarvisi* using the specific pair of primers reported in Supplementary Methods Table 1. The following list comprises the Genbank accession numbers of the 8 putative MOY orthologous protein sequences: BoMOY-A MK165746, BoMOY-B MK165747, BjMOY MK165748, BdMOY MK165749, BtMOY MK165750, BIMOY MK165751, BzMOY MK165752, RzMOY MK165753, ZcMOY MK165754.

## Methods references

- 31 Gabrieli, P. *et al.* Sex and the single embryo: early development in the Mediterranean fruit fly, *Ceratitis capitata*. *BMC Dev. Biol.* **10**, 12 (2010).
- 32 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652, doi:10.1038/nbt.1883 (2011).
- 33 Haas, B. J. *et al.* *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512, doi:10.1038/nprot.2013.084 (2013).
- 34 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).

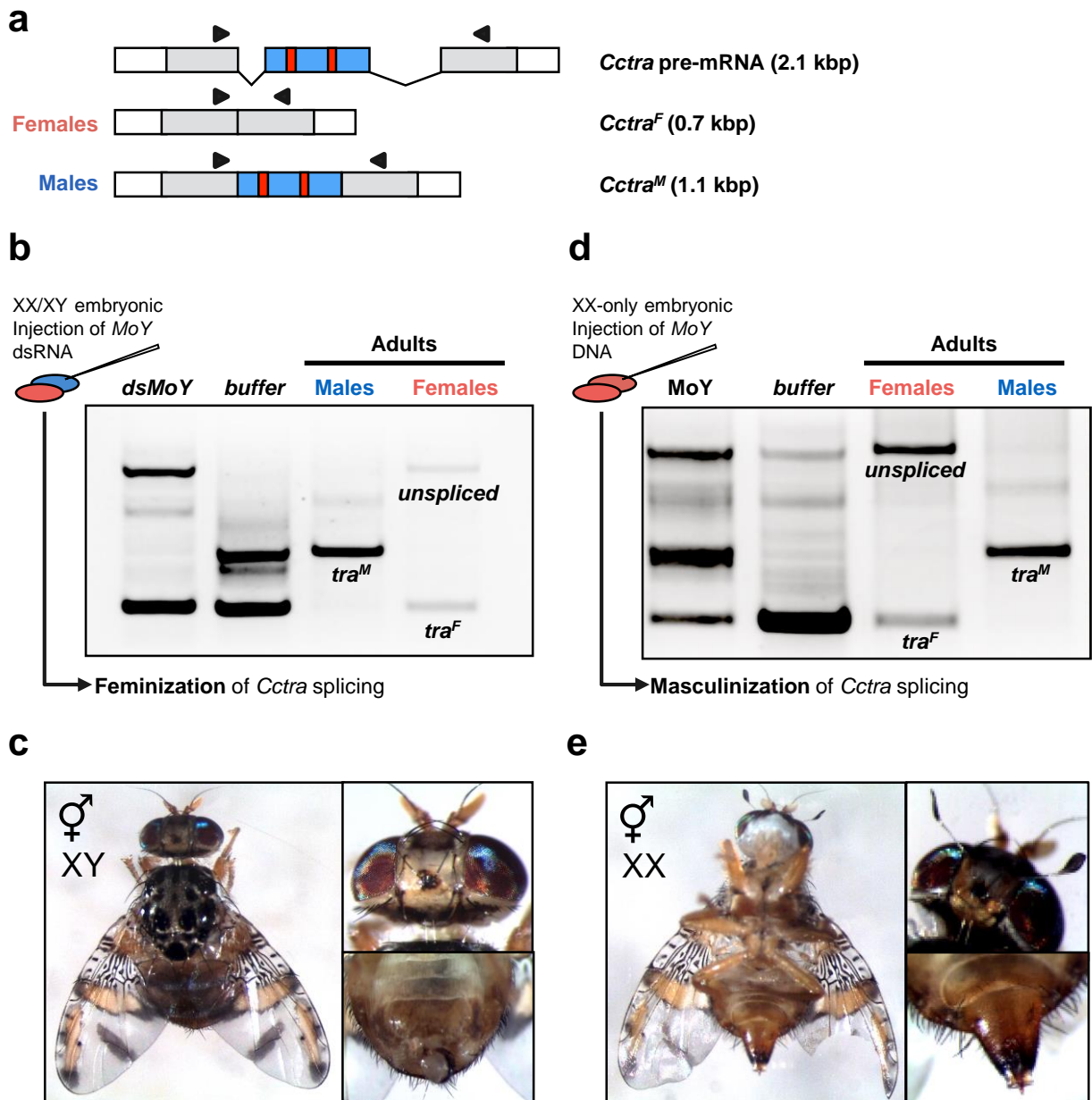
- 35 Salvemini, M. *et al.* *De novo* assembly and transcriptome analysis of the Mediterranean fruit fly *Ceratitidis capitata* early embryos. *PLoS One* **9**, e114191, doi:10.1371/journal.pone.0114191 (2014).
- 36 Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525-527, doi:10.1038/nbt.3519 (2016).
- 37 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
- 38 Anleitner, J. E. & Haymer, D. S. Y enriched and Y specific DNA sequences from the genome of the Mediterranean fruit fly, *Ceratitidis capitata*. *Chromosoma* **101**, 271-278 (1992).
- 39 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 40 Holmes, D. S. & Bonner, J. Preparation, molecular weight, base composition, and secondary structure of giant nuclear ribonucleic acid. *Biochemistry* **12**, 2330-2338 (1973).
- 41 Li, H. BFC: correcting Illumina sequencing errors. *Bioinformatics* **31**, 2885-2887, doi:10.1093/bioinformatics/btv290 (2015).
- 42 Chapman, J. A. *et al.* Meraculous: *de novo* genome assembly with short paired-end reads. *PLoS One* **6**, e23501, doi:10.1371/journal.pone.0023501 (2011).
- 43 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 44 Hackl, T., Hedrich, R., Schultz, J. & Forster, F. Proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics* **30** (21), 3004-3011 doi:10.1093/bioinformatics/btu392 (2014).
- 45 Kato, A., Albert, P. S., Vega, J. M. & Bichler, J.A. Sensitive fluorescence *in situ* hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotech. Histochem.* **81**, 71-78, doi:10.1080/10520290600643677 (2006).

- 46 Carabajal Paladino, L. Z., Nguyen, P., Šíchová, J. & Marec, F. Mapping of single-copy genes by TSA-FISH in the codling moth, *Cydia pomonella*. *BMC Genet.* **15** Suppl 2, S15, doi:10.1186/1471-2156-15-S2-S15 (2014).
- 47 Krylov, V., Tlapakova, T. & Macha, J. Localization of the single copy gene *Mdh2* on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet Genome Res* **116**, 110-112, doi:10.1159/000097427 (2007).
- 48 Sagri, E. et al. Housekeeping in Tephritid insects: the best gene choice for expression analyses in the medfly and the olive fly. *Sci. Rep.* **7**, 45634. doi:10.1038/srep45634 (2017).
- 49 Montague, T. G., Cruz, J. M., Gagnon, J. A., Church, G. M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-407, doi:10.1093/nar/gku410 (2014).



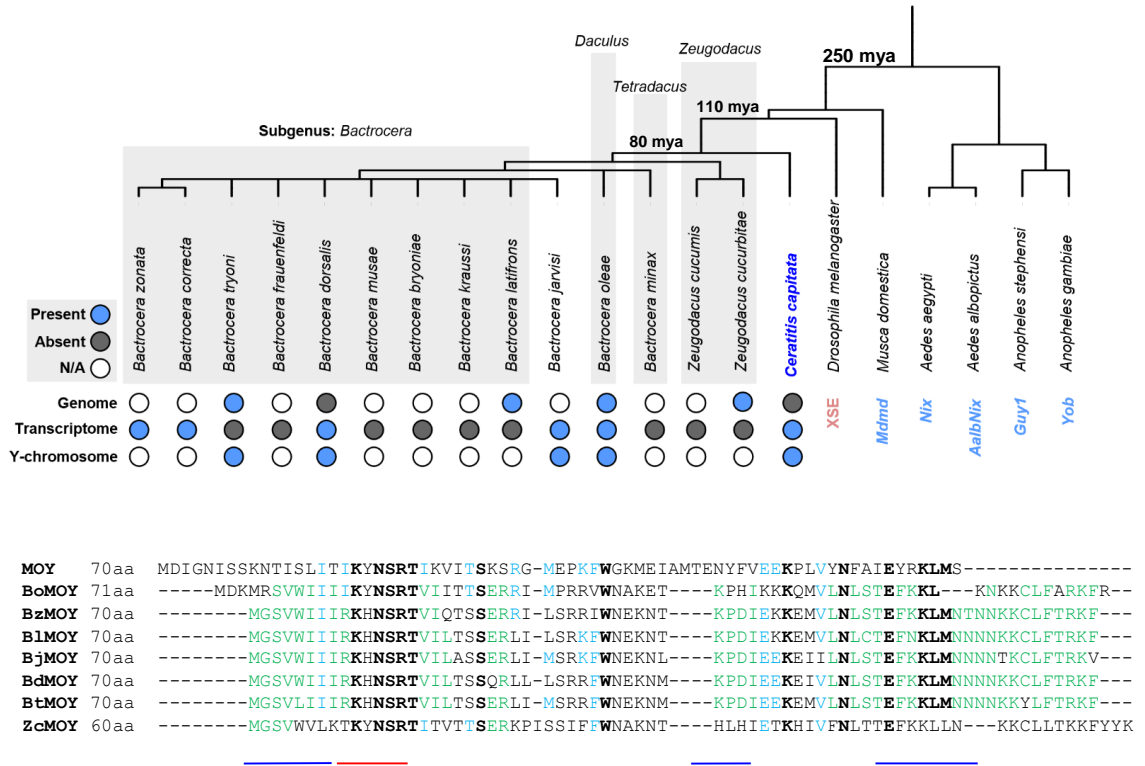
**Figure 1 | Identification of the medfly Y-linked Male determining factor (*MoY*), a single copy gene active during early embryogenesis.**

**a**, Volcano plot of the 195 differentially expressed transcripts at 4-8 h AEL, with FDR < 0.05. Pink dots indicate female-biased transcripts while azure dots indicate male-biased transcripts. Blue dots indicated transcripts with a CQ value of 0 (putative Y-linked transcripts). The transcript corresponding to *MoY* (DN40292\_c0\_g3\_i1) is indicated by arrow. Dots at the top with a black border indicate transcripts where the FDR values were lower, but were set to 1e-40 for visualisation. **b**, Fluorescence *in situ* hybridization of *MoY* (2 green signals) on mitotic chromosomes stained with DAPI (blue); twin signals (because of two sister chromatids) locate *MoY* on the long arm of the Y chromosome near the centromere. **c**, 12 Kb long genomic region (from PacBio Canu assembly; GenBank accession number: MK330842) showing in red the *MoY* transcript (8) antisense RNA (in red 9-10) and weakly related sequences (in red discontinuous lines), in violet (4) and green (5) a repeated sequence transcribed in both strands, in light brown 3 unrelated transcripts (1-3), in brown (6) and light violet (7) other transcriptional novel units. **d**, Expression profile of *MoY* during medfly embryogenesis. A peak at 15 h is observed.

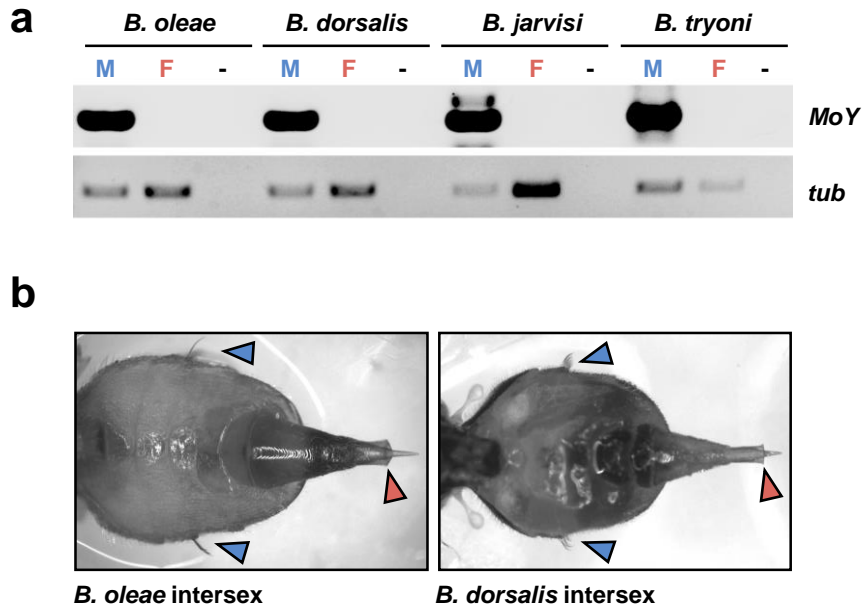


**Figure 2 | *MoY* is necessary and sufficient for male sex determination.** **a**, Scheme of *Cctra* gene and male- and female-specific transcripts. Coding region is in gray, male-specific exons are in blue, red bars indicate stop codons in the male-specific exon. Black arrows indicate primers used for RT-PCR. **b**, Transient embryonic *MoY* RNAi is sufficient to feminize XY individuals. RT-PCR analysis showing splicing patterns of *Cctra* in 8 h old embryos, following dsRNA *MoY* or buffer injections at 0-1 h AEL, compared to wild type adult flies. **c**, An example of an XY adult intersex developed from *MoY* dsRNA injected embryos (Table 1, 1#) showing male genitalia and a female-like head, lacking the orbital setae (for wild type reference see Supplementary Fig. 2). **d**, XX-only embryonic injections of genomic DNA containing *MoY* is sufficient to induce male-specific splicing of *Cctra*, but not the buffer injection, compared to wild type adult flies. **e**, An example of a XX adult intersex showing female ovipositor and a male-like head, with orbital setae.

**a**



**Figure 3 | *MoY* gene is structurally and functionally conserved in Tephritidae species. a,** Phylogenetic tree of dipteran species<sup>18</sup>, in which male determining genes or XSE (*Drosophila*) primary sex determining signals have been molecularly characterized. *MoY* orthologues were searched in 14 Tephritidae species, and found in 8 either in genome and/or transcriptome available databases at NCBI (Supplementary Information Notes 5-12) (The lack of hits in 6 out of 14 assembled transcriptomes is likely due to insufficient sequencing coverage). **b,** Manual alignment of 8 *MoY* orthologous proteins (*BcMoY* was excluded because incomplete) from *Ceratitidis capitata* (*MOY*), *Bactrocera oleae* (*BoMOY*), *B. zonata* (*BzMOY*), *B. latifrons* (*BIMOY*), *B. jarvisi* (*BjMOY*), *B. dorsalis* (*BdMOY*), *B. tryoni* (*BtMOY*), and *Zeugodacus cucurbitae* (*ZcMOY*), respectively. In black bold are indicated amino acids conserved in all 8 species, in light blue those conserved between *Ceratitidis* and some other species and in green those conserved in Tephritidae species other than *Ceratitidis*. In red, a *MoY* conserved short region, named as *MOY-R*, in blue *MoY* conserved *Bactrocera* Regions, named as *BMOY-R1*, *-R2* and *-R3*.



**Figure 4 | *MoY* gene is Y-linked in *Bactrocera* species species.** **a**, PCR on male and female genomic DNA showing Y-linkage of *MoY* orthologues in *Bactrocera oleae* (*B. oleae*), *B. dorsalis*, *B. jarvisi*, and *B. tryoni*. Positive control: beta tubulin. **b**, XY intersexes of *B. oleae* and *B. dorsalis*, following embryonic *BoMoY* and *BdMoY* RNAi, which show the presence of both female-specific (ovipositor; red arrow) and male-specific characteristics (abdominal lateral bristles; blue arrows).