Identification of palindromes in *Timema* stick insects

Jonathan Klopfenstein

Supervisor: Marc Robinson-Rechavi
Direct supervisor: Kamil Jaron

Timema poppensis perfectly camouflaged on its host, Redwood *Sequoia sempervirens*, California.
Source: BMC Ecology image competition: the winning images. *BMC Ecology* 2013

Abstract

Timema is a genus of stick insects, native to the West of the USA, that comprises both sexual and asexual species. In this genus, there is evidence of at least seven independent transitions from sexual to asexual reproduction. As asexual species are subject to Muller's ratchet, avoiding it is necessary for the persistence of asexual lineages over long time periods. Studies on the human Y chromosome, also subject to Muller's ratchet, suggest that one way to escape it involves gene conversion events driven by genomic palindromes. Collinearity analysis on two published genomes of asexual species suggests that they tend to accumulate palindromes, consistent with palindrome-driven gene conversion evidence of the human Y chromosome. In this study, we report the current counts of palindromes in sexual and asexual species of *Timema* stick insects and discuss trends on palindromes identification.

Introduction

Timema is a genus of stick insects native to the USA that comprises both sexual and asexual species. During their evolution, there were at least seven independent transition from sexual to asexual reproduction mode (Figure 1), changing the mode of reproduction to apomictic parthenogenesis, which is the production of viable eggs from female without male fertilization (Schwander, Henry and Crespi, 2011). As this mode of reproduction does not allow recombination, due to the lack of meiosis, and does not involve gamete fusion (and therefore contribution of different genetic material onto egg production) (Figure 2), obligatory parthenogenetic species are subject to Muller's ratchet, the irreversible accumulation of deleterious mutations (Felsenstein, 1974; Hillis, 2007). As genomes of asexual species are inherited as indivisible pieces, the idea behind Muller's ratchet is that once that the least mutated genome acquires a deleterious mutation, there is no expectation of finding a genome with fewer mutations within that species. Thus, in theory, asexual species tend to fix and accumulate deleterious mutations in their genomes, with no way of reversing such fixations (hence the ratchet analogy), until the species eventually goes extinct. Therefore, it has been hypothesized that ancient asexual species must avoid or slow down the mutational meltdown predicted by Muller's ratchet through several suggested mechanisms, one of which being gene conversion events (Figure 3).

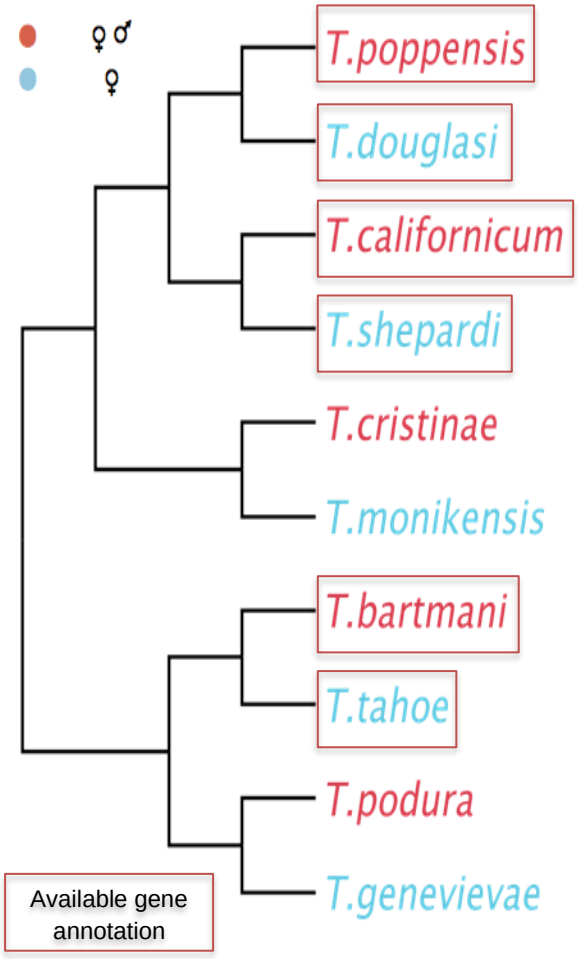


Figure 1. Cladogram of the sexual and asexual sister species of the Timema genus. The other 11 sexual species have been removed. Sexual and obligate parthenogenetic (asexual) species are depicted in red and blue, respectively. Sequenced genomes of species for which gene annotation was available are highlighted with a red square.

Evidence of abundant gene conversion in the male-specific region of the human Y chromosome shows that those events are driven through inverted blocks of collinear genes within the same chromosome (or scaffold), called palindromes (Rozen *et al.*, 2003; Bachtrog, 2013; Trombetta and Cruciani, 2017). Palindromes of the male-specific region of the human Y chromosome are large and near identical inverted repeats, typically presenting a sequence identity of 99.97% between their arms, that allows the formation of unusual chromatin structures, such as inter and intra-palindromic loops, able to optimize arm to arm gene conversion (Figure 4) (Rozen *et al.*, 2003; Bachtrog, 2013).

Genome sequencing and collinearity analysis of *Folsomia candida*, an asexual springtail living in the soil, and *Adineta vaga*, a bdelloid rotifer (also asexual), show that these species have

accumulated palindromes (Flot *et al.*, 2013; Faddeeva-Vakhrusheva *et al.*, 2017). As there is evidence that some asexual species accumulated palindromes in their genome, it is suggested that these palindromes might allow and optimize gene conversions in the same manner as those present on the human Y chromosome, and could then facilitate the long-term survival of those species. Therefore, the aim of this project is to detect and compare the palindromes in the sequenced genomes of sexual and asexual sister species of *Timema* stick insects and to evaluate the reported palindromes of *F. candida* and *A. vaga* for their ability to drive gene-conversion event as well as generally discussing the specificity of palindromes as tools for gene conversion. (Figure 1).

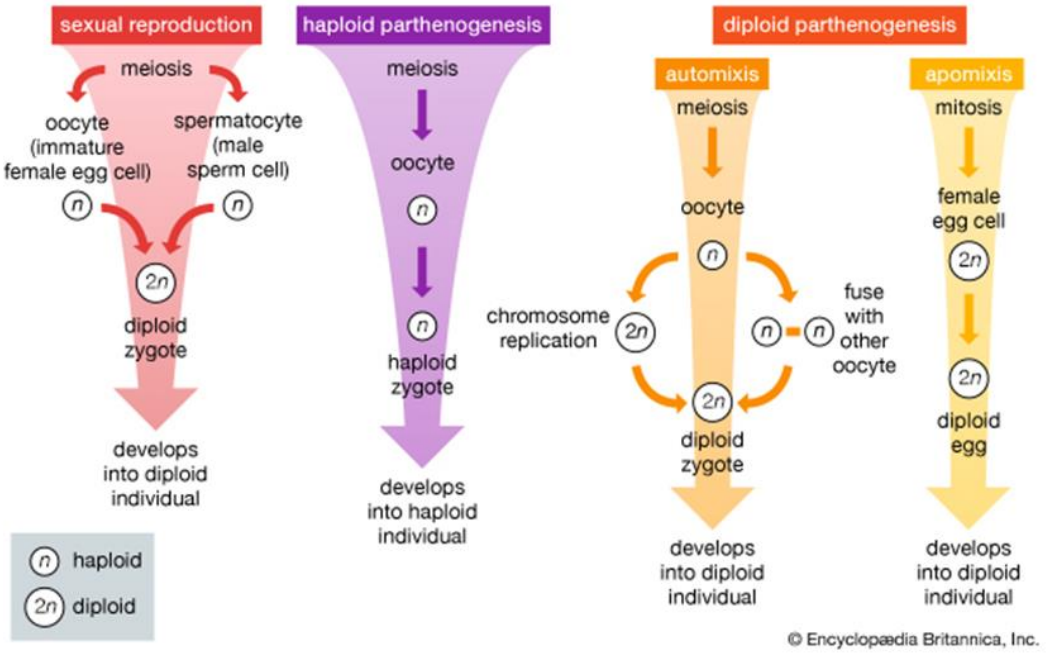


Figure 3. Schematic of the process of sexual reproduction versus several forms of parthenogenesis. Source: Encyclopaedia Britannica.

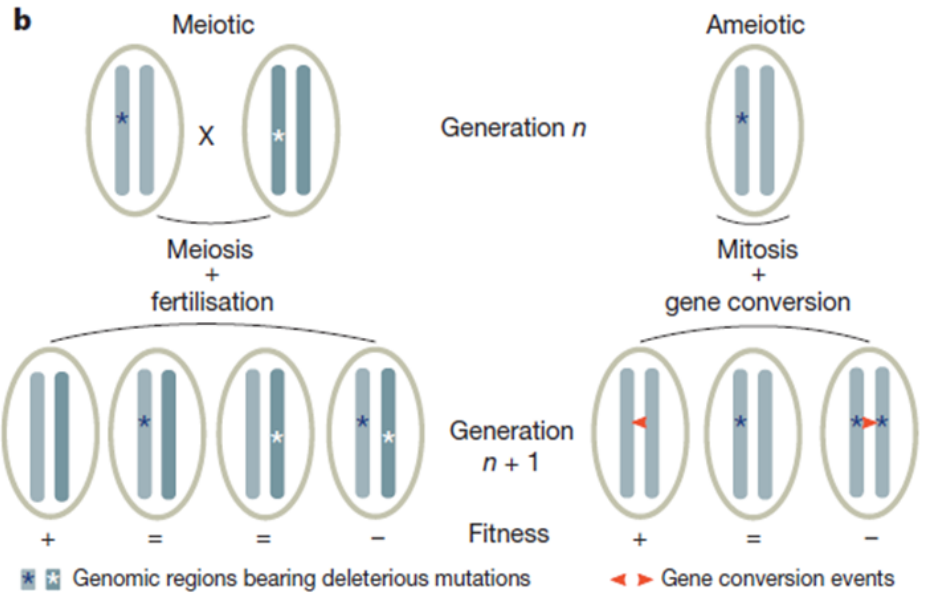


Figure 4. Gene conversion and its evolutionary consequences in ameiotic organisms compared to meiotic organisms. Source: (Flot et al.*, 2013).*

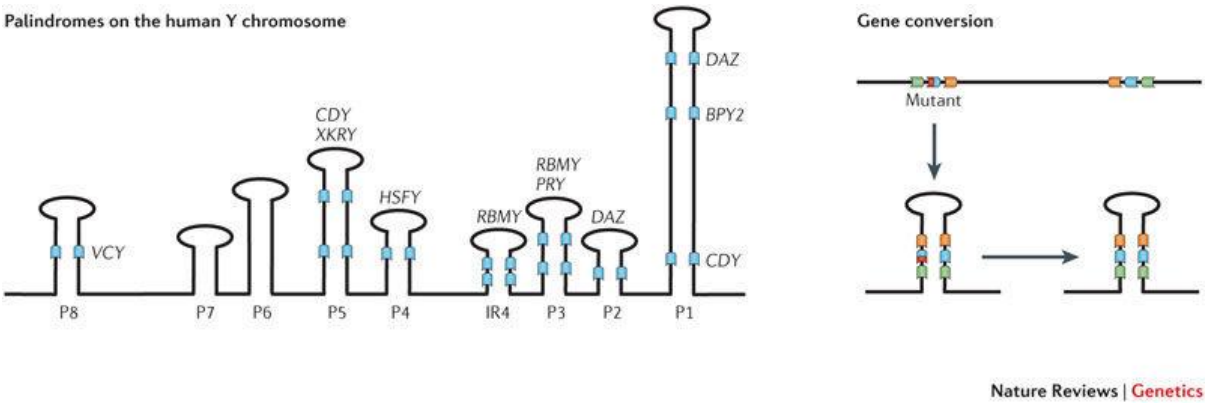


Figure 5. Palindromes on the human Y chromosome and model of palindrome-driven gene conversion. Male specific gene families are depicted in blue. Source: (Bachtrog, 2013)

Method

Collinearity analysis and palindromes identification is performed using MCScan algorithm implemented in MCScanX package (Wang *et al.*, 2012). MCScan is an algorithm able to detect synteny and collinearity, i.e. collinear blocks of genes, among genomes. Given a gene annotation file and the BlastP results of all proteins against themselves, the algorithm tries to build chains of matching gene pairs using the gene annotation file as reference for the position of the genes (Figure 5). Palindromes are a subset of collinear blocks that are localized within the same scaffold in an inverted manner. Their detection is conducted by parsing the output of MCScanX.

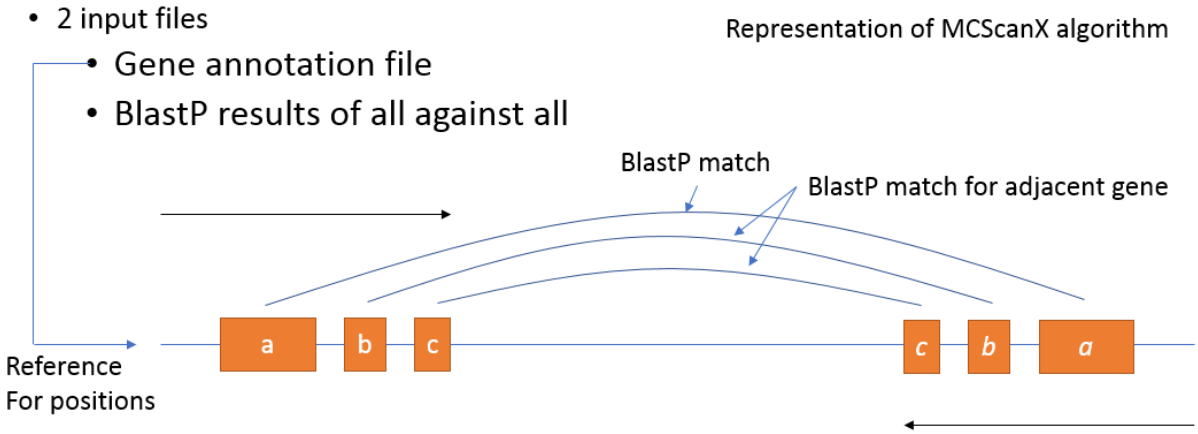


Figure 6. Schematic representation of MCScan algorithm on a palindrome. Genes are represented as orange blocks and their inverted counterparts, linked by a curve, are represented using italic letters.

Such genomes structures, as represented in Figure 5, are taught to be able to optimize gene conversion events, under the condition that the arms of the palindromes are nearly identical.
BlastP search of all against all, required by MCScanX, was performed using a E-value cut-off of 1e-10 (Altschul *et al.*, 1990).
All scripts and code used to generate this study are available in the dedicated Github repository for this study: https://github.com/AsexGenomeEvol/palindromes.

Results and discussion

Palindromes identification was performed for the *Timema* genomes of which gene annotation was available: *T. poppensis, T. douglasi, T. californicum, T. shepardi,T. bartmani* and *T. tahoe* (Figure 1).

Table 1. Results of collinearity analysis on Timema stick insects sexual and asexual sister species. Sexual species and asexual species are depicted in red and blue, respectively.

Counts of collinear blocks and palindromes per analysed *Timema* genomes using default parameters.

Species	Number of collinear blocks	Number of palindromes
T.poppensis	0	0
T.douglasi	0	0
T.californicum	0	0
T.shepardi	0	0
T.bartmani	0	0
T.tahoe	0	0

We found no collinear blocks, and thus no palindromes, using MCScanX default parameters in any of the six analysed species (Table 1). Parameters were chosen to be comparable to the collinearity analysis of *F. candida* and *A. vaga*, asserted by personal communication with the authors (Flot *et al.*, 2013; Faddeeva-Vakhrusheva *et al.*, 2017). The results were unexpected, and led us to an examination of MCScanX parameters.

The package has several parameters, of which three of them are of particular interest:

- MATCH_SIZE: the number of genes required to call a collinear block, set by default to 5.
- MAX_GAPS: the number of gaps (non-matching gene pairs) allowed within a collinear block, set by default to 25.
- E_VALUE: the metric generated by MCScanX for collinear block significance, set by default to 1e-05.

The default parameters and their usage in the collinearity analysis of *F. candida* and *A. vaga* invited to assess the biological relevance of palindromes calling regarding gene conversion. First, as the genes involved are inverted repeats in the call of a palindrome, the minimum number of genes required for the call can be decreased as to detect smaller palindromes, that still could serve as templates for gene conversion. Secondly, the presence of gaps in the call of a palindrome, in the context of gene conversion, was very questionable. Indeed, using the structure of the palindromes of the male specific region of the human Y chromosome as the standard for gene conversion within palindromes, the arms of the palindromes are required to be nearly identical (Rozen *et al.*, 2003). The fact that gaps are allowed does not concord with this standard, and should be reduced to a minimum, if not completely absent.

Therefore, a parameter exploration was performed, using the genomes of the springtail *F. candida* and the rotifer *A. vaga*, to assess the biological relevance of palindromes calling. The

results, using different values of matching size and maximum allowed gaps, are presented in Table 2.

Table 2. Number of detected palindromes in the genomes of the springtail F. candida and rotifer A. vaga when varying the parameters of maximum allowed gaps (max gap) and minimum number of matching genes required to call a collinear block (match size). The reported results, generated by using default parameters of MCScanX, of each paper are highlighted in red.

Counts of detected palindromes in genome of springtail *F. candida* and rotifer *A. vaga* upon parameters exploration.

		Number of detected palindromes	
max gap	**match size**	**Springtail**	**Rotifer**
Default: 25	**Default: 5**	**11**	**17**
0	1	3	12
0	2	3	12
0	3	3	12
0	4	3	12
0	5	3	12
1	1	5	27
1	2	5	27
1	3	5	27
1	4	5	27
1	5	4	24
5	1	4	21
5	2	7	21
5	3	7	21
5	4	8	21
5	5	6	18

Overall, altering the parameters gives very different results. When considering detection with no gaps allowed within the arms of the palindromes, their number in the springtail genome becomes significantly reduced (from 11 to 3). Additionally, when the maximum allowed gaps value is set to 5, increasing the match size should decrease the number of detected palindromes, but for the springtail, the number of called palindromes increases, which is counter-intuitive. This is explained by the E-value cut-off (set to the default 1e-05), which filters palindromes of few genes due to low significance. If we remove this cut-off, by setting it to 10 for example and thus not filtering the calls, we observe the expected decrease in the number of detected palindromes (Supplementary Figure 1). As for the rotifer genome, the number of detected palindromes, when using the most stringent parameters (i.e. no gaps allowed), is also reduced (17 to 12), but there is still an appreciable amount of them. Therefore, the fact that the majority of the palindromes remains after trimming for their potentially biological relevance in gene conversion suggests that they could be of importance and that this species might undergo gene conversion through palindromes.

Interestingly, two of the three potentially relevant palindromes of the springtail genome are organized in a unique fashion (Figure 6). These two palindromes have a shared arm, and the two other arms are adjacent to each other. However, details about this particular gene

structure are not known yet. In the rotifer reference article, the contents of the palindromes are not discussed, and the content of these observed “true” palindromes, regarding their potential to drive gene conversion, is thus not discussed in this project.

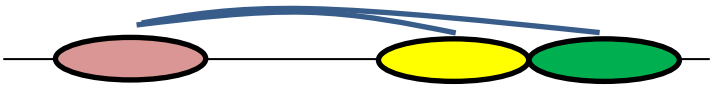


Figure 6. Schematic representation of the structure of two of the three potentially relevant palindromes detected in the springtail genome.

Returning to *Timema* species, after verifying that the absence of collinearity isn’t due to technical errors, we do observe calls of collinear blocks and palindromes if the E-value cut-off is removed (i.e. set to 10) (Table 3).

Table 3. MCScanX outputs on Timema sexual and asexual species. Sexual and asexual species are depicted in red and blue, respectively. Sexual and asexual sister species are paired together within each block of rows.

Counts of collinear blocks and palindromes per analysed *Timema* species, E-value = 10 and MAX_GAPS = 25.

match_size	♂♀ Species	Number of collinear blocks	Number of palindromes	♀ Species	Number of collinear blocks	Number of palindromes
1	*T.poppensis*	78	5	*T.douglasi*	113	2
2		41	3		61	2
3		1	0		1	0
4		0	0		0	0
5		0	0		0	0
1	*T.californicum*	124	7	*T.shepardi*	112	5
2		66	4		70	1
3		8	0		1	0
4		1	0		1	0
5		0	0		0	0
1	*T.bartmani*	101	3	*T.tahoe*	133	6
2		46	1		69	4
3		0	0		2	0
4		0	0		0	0
5		0	0		0	0

However, since the conditions on calling were stretched out to a maximum, the biological relevance of these collinear blocks and palindromes calls cannot be trusted. Those results might only serve here as a control to reassure that the analysed *Timema* genomes must certainly contain some collinearity and palindromes, compared to the complete absence of calls when using default parameters.

Therefore, the absence of observations in terms of collinearity, and consequently palindromes (Table 1), is most probably due to technical factors on the assembly and annotations of the genomes. While the annotation of the genomes is most likely incomplete, the current status of the assembly is very fragmented (Supplementary Table 1), and thus, even if collinear blocks were detected, the calling of palindromes would be limited, as they might be assigned to different scaffolds and consequently would not be detectable.

Conclusions

Sexual and asexual *Timema* species do not currently show evidence of palindromes, nor collinearity. The absence of calls is most probably an artefact of the current assembly and annotation status. Therefore, the reported results in this study are not definitive, and as the assembly and annotation of those genomes improves, very different results, such as actual counts of significant collinear blocks and palindromes, are expected. Indeed, the fact that these species present absolutely no collinearity in their genomes is unlikely.

An important result is that palindromes calling using MCScanX, which is currently the most used package in the world to detect synteny and collinearity among genomes, is heavily dependent on the arbitrary choices of parameters. Furthermore, the current default parameters, and the use of MCScanX without specifying parameters, are probably not the best choice for gene-conversion-allowing palindromes identification. In general, palindrome identification for their potential role in gene conversion requires a more robust standardization.

Methods for detecting palindromes involved with gene conversion should intuitively exclude the presence of gaps within the arms of the palindromes, but the exact values for standardization should be performed by conducting a meta-analysis of all genomes for which evidence of palindrome-driven gene conversion exists. Following that analysis, condition required for a palindrome to be able to drive gene conversion, such as number of genes, total length of sequence, sequence similarity and range of number of gaps allowed, could be assessed and used as standard for the analysis and reanalysis of new and already published genomes.

Bibliography

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Bachtrog, D. (2013) 'Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration', *Nature Reviews Genetics*. Nature Publishing Group, 14(2), pp. 113–124. doi: 10.1038/nrg3366.

Faddeeva-Vakhrusheva, A. *et al.* (2017) 'Coping with living in the soil: the genome of the parthenogenetic springtail Folsomia candida', *BMC genomics*. BMC Genomics, 18(493), pp. 1–14. doi: 10.1186/s12864-017-3852-x.

Felsenstein, J. (1974) 'The evolutionary advantage of recombination.', *Genetics*, 78(2), pp. 737–56. Available at: http://www.ncbi.nlm.nih.gov/pubmed/4448362 (Accessed: 22 December 2017).

Flot, J.-F. *et al.* (2013) 'Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga', *Nature*, 500(7463), pp. 453–457. doi: 10.1038/nature12326.

Hillis, D. M. (2007) 'Asexual Evolution: Can Species Exist without Sex?', *Current Biology*. Cell Press, 17(14), pp. R543–R544. doi: 10.1016/J.CUB.2007.05.015.

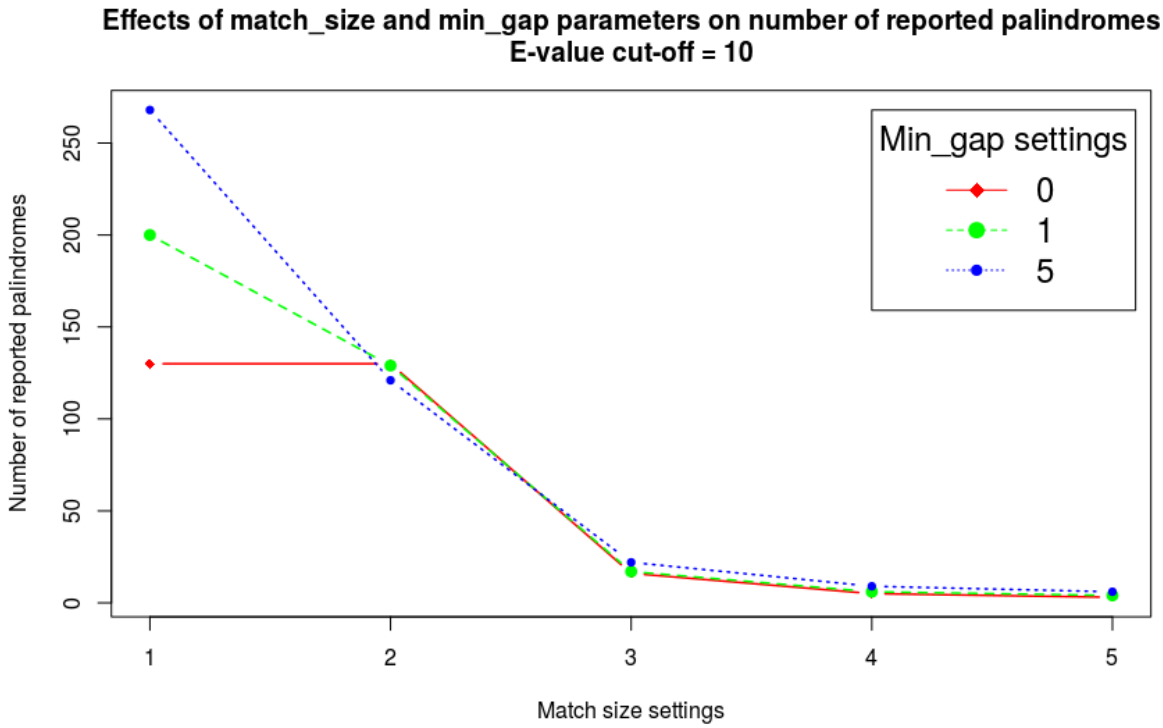
Rozen, S. *et al.* (2003) 'Abundant gene conversion between arms of palindromes in human and ape Y chromosomes', *Nature*. Nature Publishing Group, 423(6942), pp. 873–876. doi: 10.1038/nature01723.

Schwander, T., Henry, L. and Crespi, B. J. (2011) 'Molecular evidence for ancient asexuality in timema stick insects.', *Current biology : CB*. Elsevier, 21(13), pp. 1129–34. doi: 10.1016/j.cub.2011.05.026.

Trombetta, B. and Cruciani, F. (2017) 'Y chromosome palindromes and gene conversion', *Human Genetics*. Springer Berlin Heidelberg, 136(5), pp. 605–619. doi: 10.1007/s00439-017-1777-8.

Wang, Y. *et al.* (2012) 'MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity', *Nucleic Acids Research*, 40(7), pp. e49–e49. doi: 10.1093/nar/gkr1293.

Appendix



Supplementary Figure 1. Effects of match_size and min_gap parameters on the number of reported palindromes without E-value filtering. The expected decrease of number of palindromes is observed, due to the increasing limitation on palindrome calling.

Supplementary Table 1.

Counts of scaffolds in genome assembly per species.

Number of scaffolds	Species
10172	*T. poppensis*
7850	*T. douglasi*
8686	*T. californicum*
7933	*T. shepardi*
6988	*T. bartmani*
6041	*T. tahoe*
133	*F. candida* (springtail)
3091	*A. vaga* (rotifer)

Acknowledgments

I wish to thank Kamil Jaron, Prof. Tanja Schwander and Prof. Marc Robinson-Rechavi for their supervision, precious advices and overall sympathy.