Supplemental File 2:

Practical guide on using YGS

What follows is practical summary, adding also more recent data. Readers should consult the original paper [35] for in depth discussion of the YGS method, including its pitfalls.

A) Assembly quality

Well assembled genomes usually yield clean YGS results (see also the next section). Assembly fragmentation produce many small scaffolds that can be difficult to reliably classify as Y / not-Y.

Another source of assembly problems is polymorphism in the sample. Samples with high genetic variation yield poorer assemblies, which translates into a lot of overlap in YGS. Ideally you should use an inbred strain, obtained by brother-sister mating for up to 20 generations. If this is not feasible (as is the case of *L. longipalpis s.l.*), individuals from colony are preferable.

B) Choice of the cut-off for candidate Y-linked scaffolds.

One point that can be tricky is the choice of the cut-off for candidate Y-linked scaffolds (we used 70% unmatched *k-mer*). Well assembled genomes usually yields two sharply separated peaks in YGS (Y and X+A), with little or no overlap (see Figs 2A and 3A in [35]); in this case the proper cut-off is self-evident. However, a less optimal assembly of the same species can have a significant amount of overlap (Fig S14A in [35]), and in these cases the choice of cut-off point should be based on the trade-off between false-positives and false-negatives. The use of a cut-off of 50% instead of 70% we would very likely get some additional real Y-linked sequences, at the cost of including some or many autosomal or X-linked sequences. Since the

main objective of the present paper was to identify the sex-determination system, we opted for a fairly stringent cut-off. If the purpose was to identify most of the Y-linked sequences a lower cut-of (always followed by PCR confirmation) would be more appropriate.