# Supplementary Figures

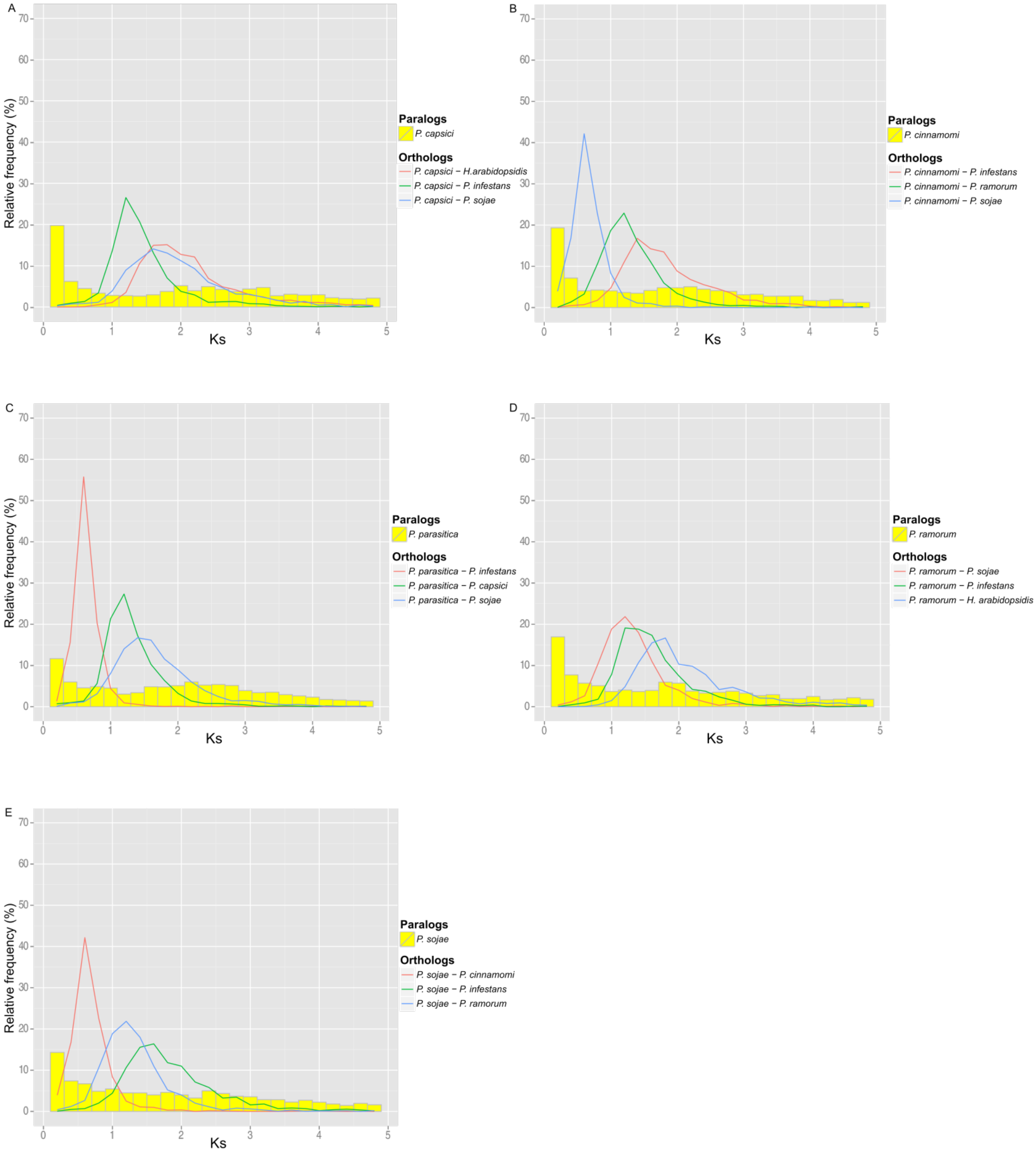


**Figure S1.** $K_s$ distributions of the full paranomes of all *Phytophthora* species included in this work and of a selection of orthologous gene pairs.

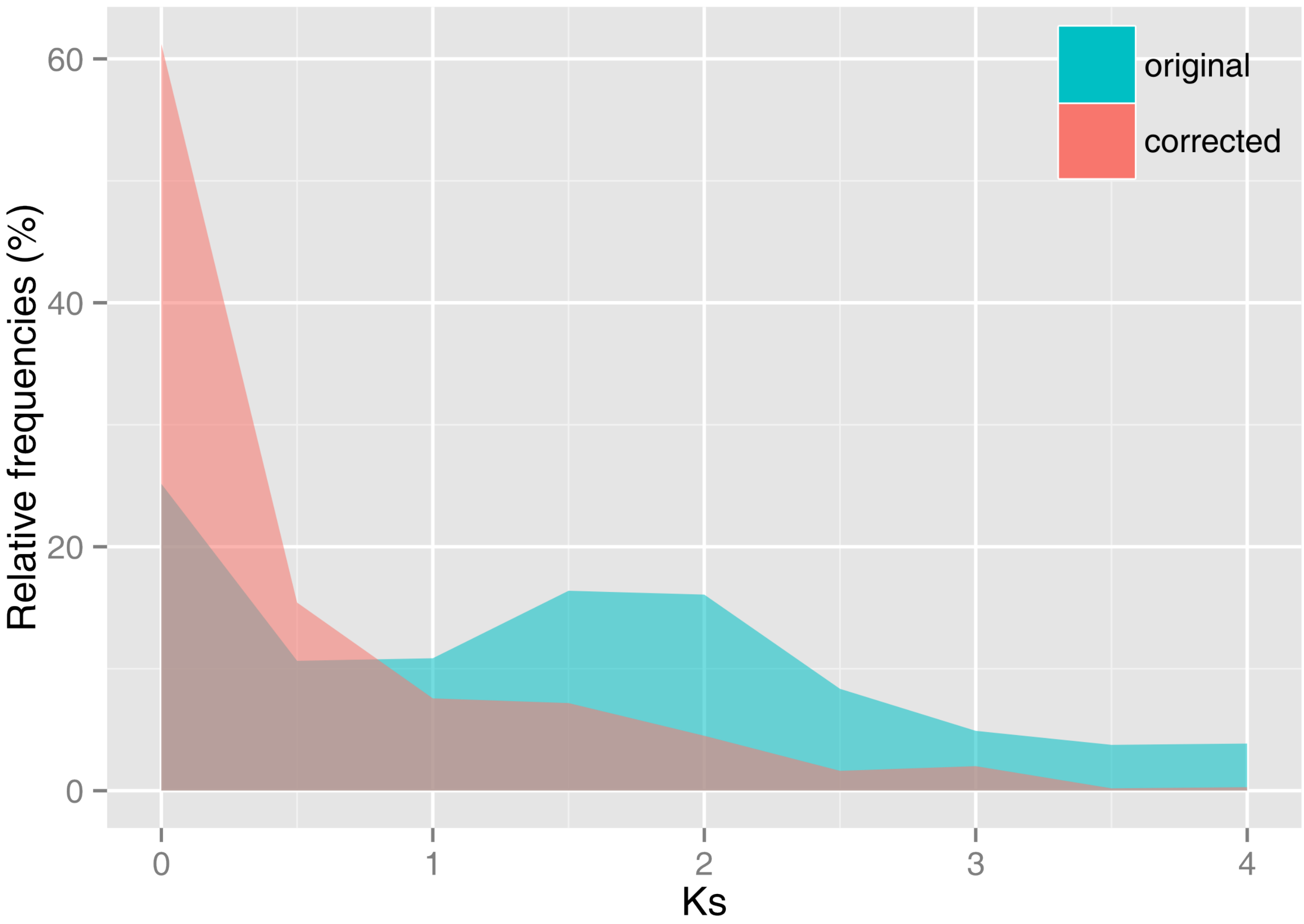


**Figure S2.** $K_s$ distributions of the *P. infestans* 2HOM paralogs derived from the original set (Martens and Van de Peer 2010). We observed inconsistencies in the original data, i.e. wrong family assignments within 2HOM blocks. Once corrected, both distributions differ significantly. Please note that for both sets the $K_s$ values constituting the distributions are calculated by Martens and Van de Peer (2010), diminishing the possibility that differences between the distributions are caused by the method of $K_s$ calculation.

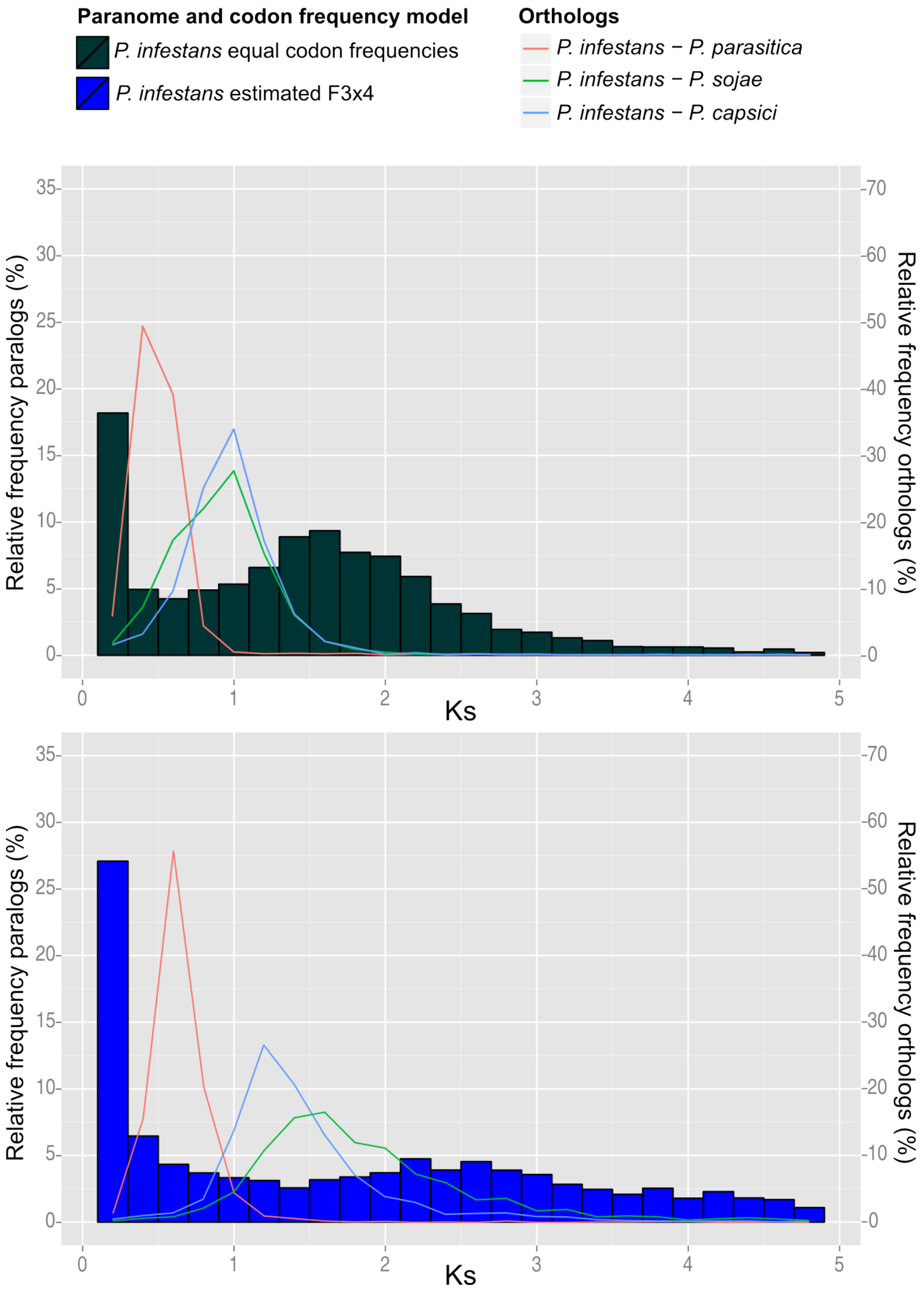


**Figure S3. $K_s$ distributions of the *P. infestans* paranome and orthologs using the equal model for the equilibrium codon frequencies (upper panel) and the F3x4 model (lower panel, same as Figure 3B)**

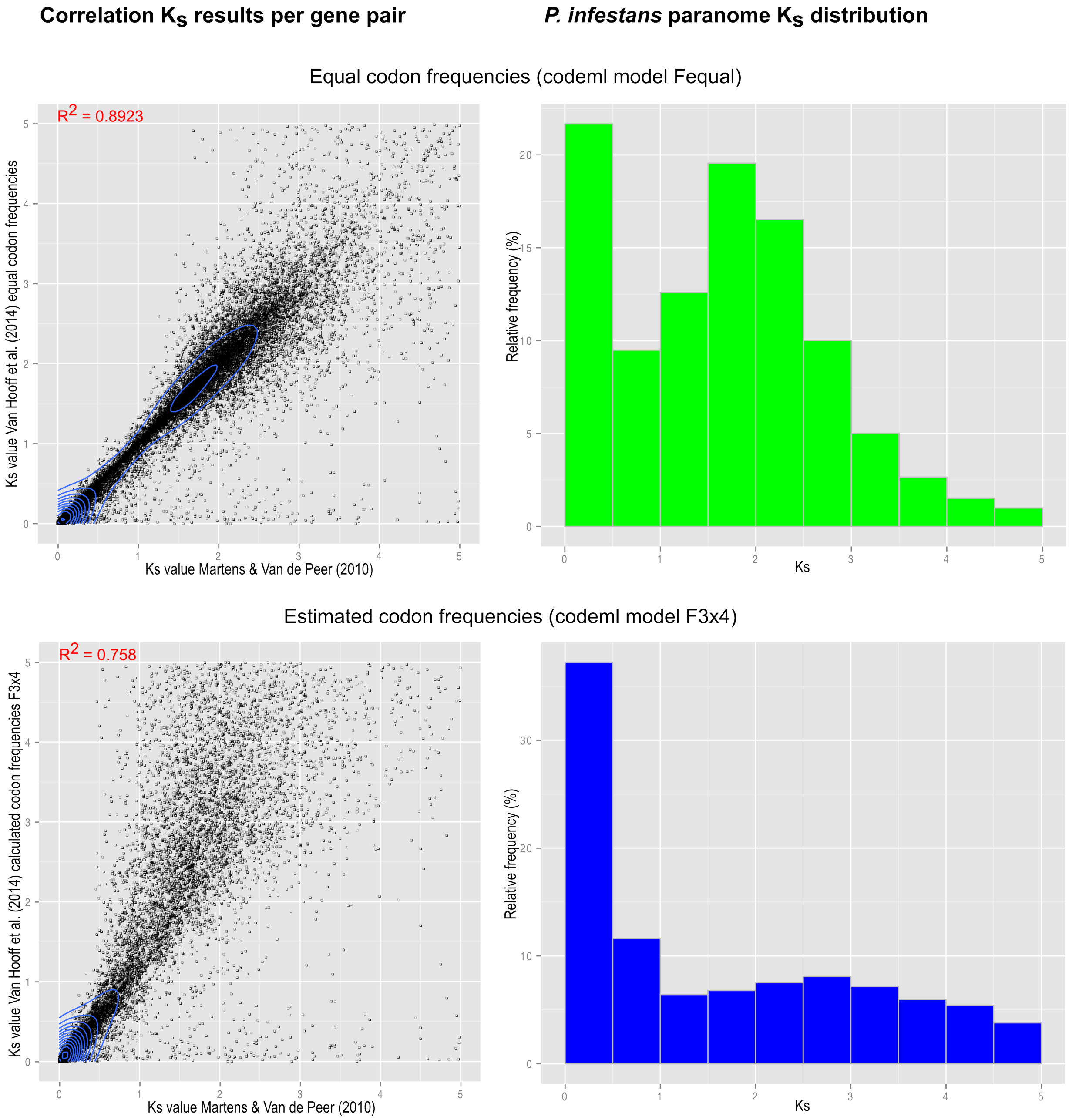
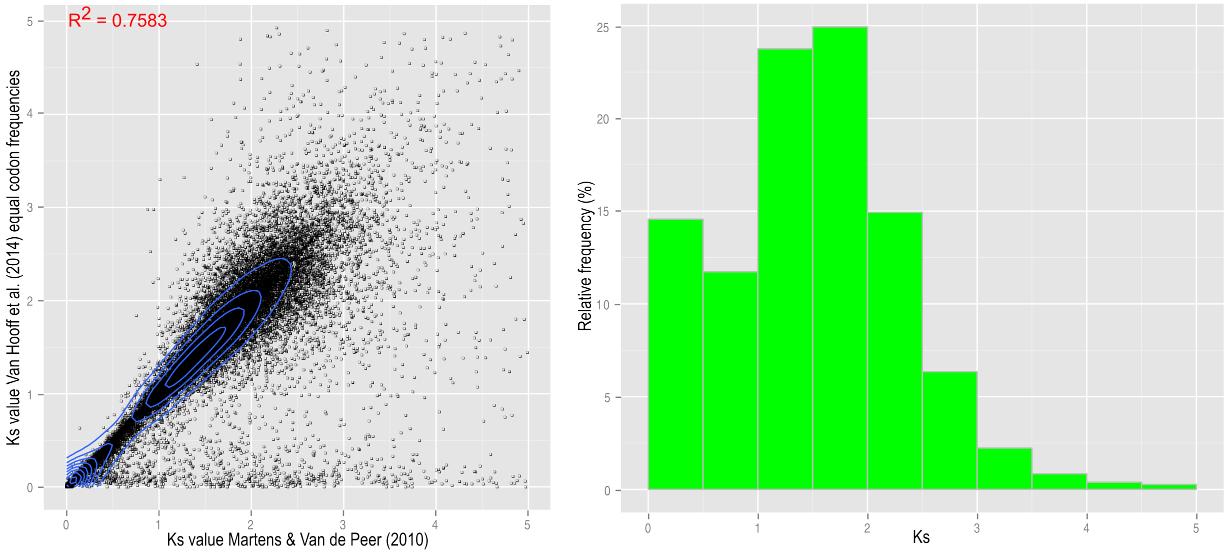


**Figure S4.** $K_s$ calculations using the equal model for the equilibrium codon frequencies (upper panels), and the F3x4 model (lower panels) applied to the *P. infestans* paralogs identified by Martens and Van de Peer (2010). The left-sided panels display scatter plots for $K_s$ values obtained by our calculation with respect to those obtained by Martens and colleagues, the blue lines indicate an increasing density of data points. Some variation between these $K_s$ values is expected due to small differences in the methods, for example in aligning the sequences. The right-sided panels contain the distributions resulting from the calculations performed by us (located on the y-axis of the left panels). Under the equal model, the $K_s$ values of Martens and Van de Peer (2010) are much better predicted than under the F3x4 model. Also, the distribution under the equal model is highly similar to Figure 4A in Martens and Van de Peer (Martens and Van de Peer 2010), whereas the distribution under the F3x4 model is comparable to Figure 3B of the current work.

**Correlation $K_S$ results per gene pair** **_P. ramorum_ paranome $K_S$ distribution**

Equal codon frequencies (codeml model Fequal)



Estimated codon frequencies (codeml model F3x4)

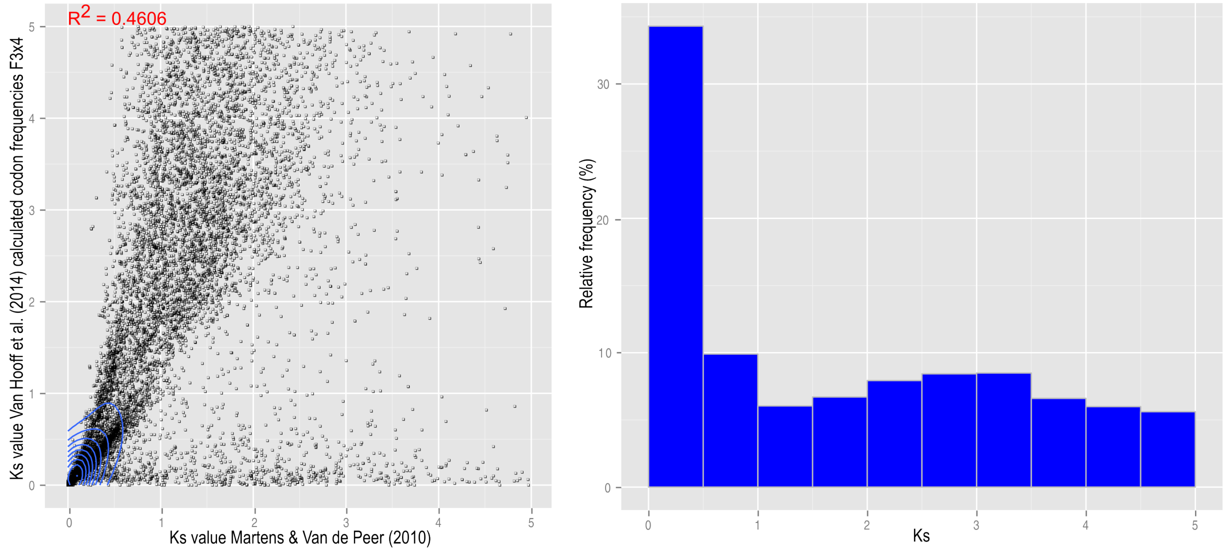
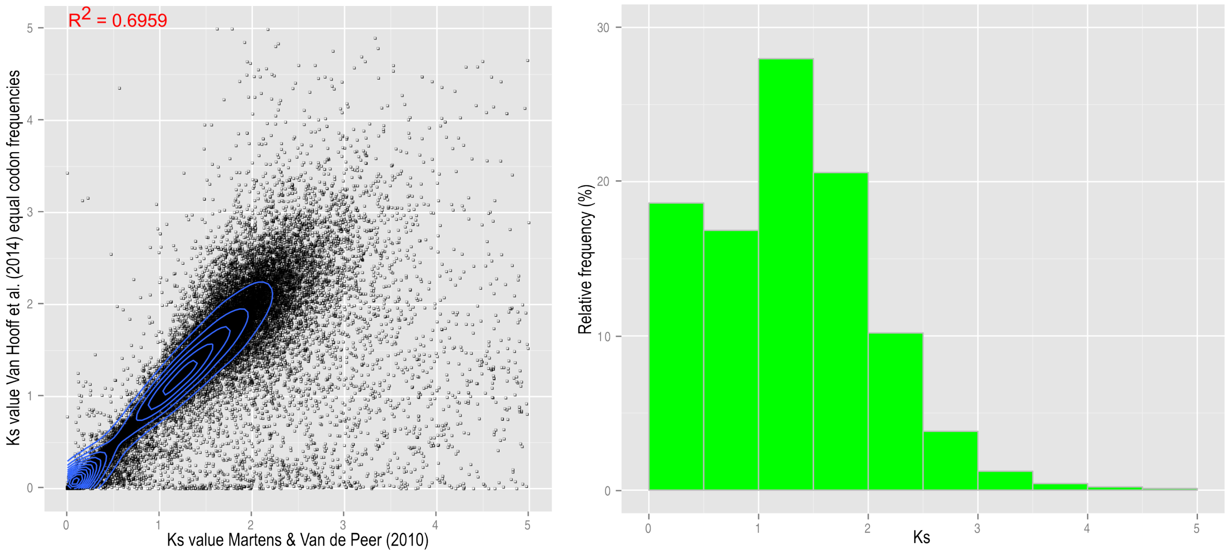


**Figure S5. Similar to Figure S4, for _P. ramorum_ paralogs. $K_s$ calculations using the equal model for the equilibrium codon frequencies (upper panels), and the F3x4 model (lower panels) applied to the _P. ramorum_ paralogs of Martens and Van de Peer (2010).**

**Correlation $K_S$ results per gene pair** **P. sojae paranome $K_S$ distribution**

Equal codon frequencies (codeml model Fequal)



Estimated codon frequencies (codeml model F3x4)

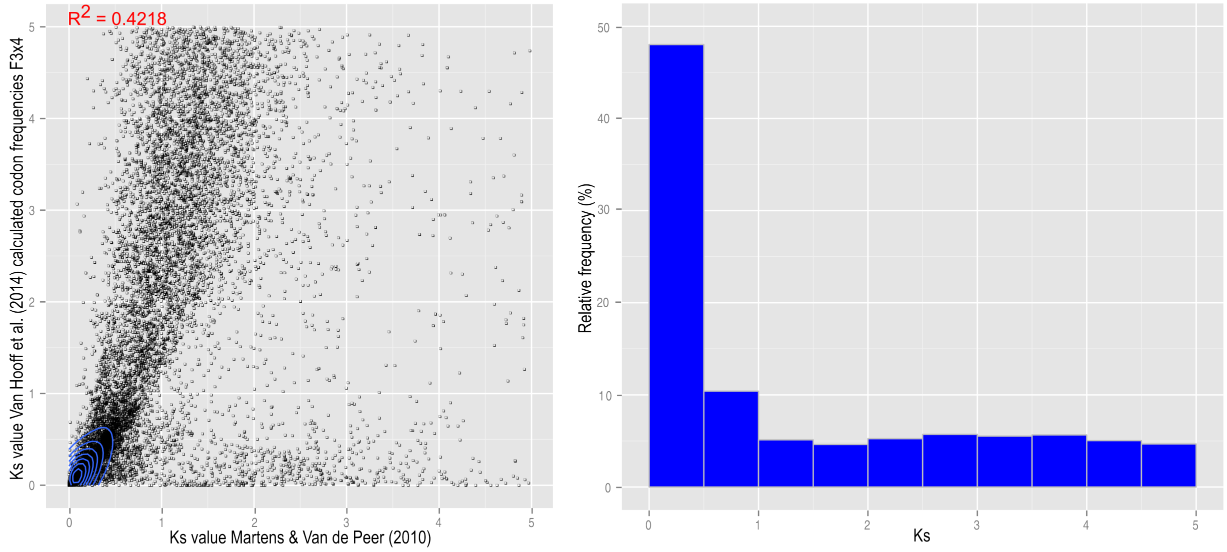


**Figure S6. Similar to Figure S4, for *P. sojae* paralogs. $K_s$ calculations using the equal model for the equilibrium codon frequencies (upper panels), and the F3x4 model (lower panels) applied to the *P. sojae* paralogs of Martens and Van de Peer (2010).**

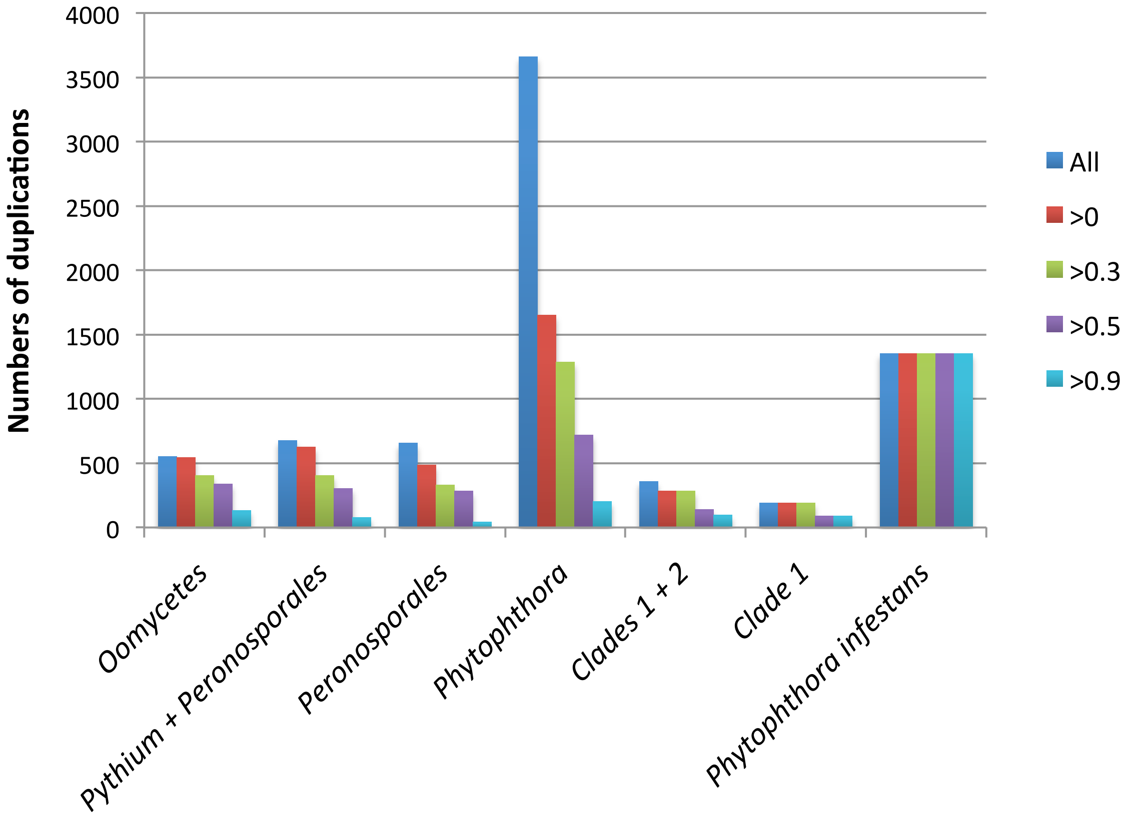


**Figure S7.** Numbers of gene duplications in *P. infestans* and its ancestors (see also Figure 1 for the species tree) as inferred from large-scale phylogenomic analysis. For each duplication the consistency score was calculated, i.e. the ratio between the number of species that still possess both gene copies and the total number of species under a duplication node. The results were subjected to increasing cut-offs, as depicted in the figure. If the score is zero, none of the today-living species has both duplicate genes and the inferred duplication is likely to be an artifact of incorrect tree topologies (Vilella et al. 2009). We consider duplications having a consistency score >0 reliable.

## References

Martens C, Van de Peer Y. 2010. The hidden duplication past of the plant pathogen phytophthora and its consequences for infection. BMC Genomics 11:353.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R et al. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Research 19:327-335.