

Testing for “Snowballing” Hybrid Incompatibilities in *Solanum*: Impact of Ancestral Polymorphism and Divergence Estimates

Thomas Städler,* Ana Marcela Florez-Rueda, and Margot Paris

Plant Ecological Genetics, Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland

*Corresponding author: E-mail: thomas.staedler@env.ethz.ch.

Associate editor: Jody Hey

Abstract

Two recent high-profile studies offered empirical evidence for a “snowballing” accumulation of postzygotic incompatibilities in *Drosophila* and *Solanum* (tomatoes). Here we present a reanalysis of the *Solanum* data that is motivated by population genetic principles. Specifically, the high levels of intraspecific nucleotide polymorphism in wild tomato species and presumably large effective population size throughout the divergence history of this clade imply that ancestral polymorphism should be taken into account when evaluating sequence divergence between species. Based on our reanalyses of synonymous-site divergence between the four focal *Solanum* species and a wide range of ancestral polymorphism, we assessed under which conditions the reported accumulation of seed sterility factors supports the snowball effect. Our results highlight the pivotal impact of levels of ancestral polymorphism and alternate divergence values, and they illustrate that robust tests of the snowball effect in *Solanum* require genome-wide estimates of divergence.

Key words: *Solanum*, snowball effect, ancestral polymorphism, molecular divergence.

Grounded in the principles of coalescent theory, evolutionary analyses of sequence data sampled within and among closely related species are of increasing importance in population genetics and molecular evolution (e.g., Wakeley 2009; Hey 2010). A basic principle is the conceptual and empirical distinction between species divergence and sequence divergence as the most recent common ancestor of a set of sequences sampled from two sister species must necessarily be older than the species’ divergence time. This is because of within-species nucleotide polymorphism at the time of lineage divergence, reflecting its long-term effective population size N_e (e.g., Li 1977; Gillespie 1991; Peterson and Masel 2009). A second, conceptually rich field in evolutionary biology is speciation genetics, whose leading explanatory framework for the evolution of postzygotic reproductive isolation is likewise built on population genetic principles (Dobzhansky–Muller incompatibilities; Coyne and Orr 2004). This model posits negative epistatic interactions among two or more loci in hybrids and thus predicts an accelerating pace of accumulating incompatibilities over time, the so-called “snowball effect” (Orr 1995; Orr and Turelli 2001).

Two recent studies independently concluded that the temporal accumulation of incompatibilities obeys the predicted snowball effect in *Drosophila* (Matute et al. 2010) and between *Solanum* species (Moyle and Nakazato 2010). More specifically, the latter study on tomatoes found that the accumulation of quantitative trait loci (QTL) for seed sterility between species was significantly nonlinear with time since divergence, whereas the accumulation of pollen sterility, fruit shape, and seed size QTL was consistent with linear. Both studies used estimates of interspecific divergence

at synonymous sites, K_s , as a proxy for time since species divergence and forced all regression models through the origin ($K_s = 0$). Even though molecular “divergence” between closely related species must comprise components of both ancestral polymorphism and subsequent interspecific divergence, this treatment in effect disallows any contribution of ancestral polymorphism.

Here we argue that this confounding effect is substantial for the *Solanum* system and analyze its potential impact under alternative divergence estimates. First, available nucleotide sequence data on extant wild tomato species show abundant within-species polymorphism at synonymous sites (π_s), ranging from 0.0106 in *Solanum habrochaites* to a high of 0.0310 in *S. peruvianum*, with *S. arcanum* (0.0188) and *S. chilense* (0.0259) showing intermediate values of π_s (calculated from sequence data in Roselius et al. 2005; Städler et al. 2005; Arunyawat et al. 2007; Tellier et al. 2011). As there are no signatures of marked demographic expansion in the sequence data for *S. chilense* and *S. habrochaites* (Städler et al. 2008; Tellier et al. 2011), these observations are consistent with ancestral π_s being on the order of ≥ 0.010 per site.

Our second more indirect line of reasoning considers effective population size at deeper time scales. The family Solanaceae is characterized by an ancestral self-incompatibility (SI) system enforcing obligate outcrossing, whereas the frequent evolutionary transitions to self-compatibility appear to be irreversible, that is, SI likely has never been regained (Igic et al. 2006; Goldberg et al. 2010). Moreover, the trans-specific distribution of the many allelic lineages segregating at the *S-RNase* gene that mediates the SI response has been used to infer ancient bottlenecks for two genera but

Table 1. Estimates of Molecular Divergence in Four *Solanum* Interspecific Comparisons.

Species pair	Weighted K_s (this study)	95% CI of K_s (this study)	Previous K_s (Moyle and Nakazato 2010)	95% CI of K_s (Moyle and Nakazato 2010)
<i>S. lycopersicum</i> – <i>S. pennellii</i> (SL–SP)	0.0267	0.0189–0.0365	0.0227	0.0157–0.0291
<i>S. lycopersicum</i> – <i>S. habrochaites</i> (SL–SH)	0.0340	0.0264–0.0438	0.0329	0.0178–0.0497
<i>S. lycopersicum</i> – <i>S. lycopersicoides</i> (SL–Slyc)	0.0544	0.0458–0.0640	0.0663	0.0353–0.1175
Tomato clade– <i>S. lycopersicoides</i> (TC–Slyc)	0.0541	0.0460–0.0638	n.a.	n.a.

NOTE.— K_s is synonymous-site divergence between species. Tomato clade (TC) refers to the pooling of sequences from several tomato species (*Solanum* section *Lycopersicon*; see text). Data were obtained from 21 nuclear loci with partial exon sequences (>31 synonymous sites; [supplementary table S1, Supplementary Material](#) online). 95% confidence intervals (CI) around the weighted K_s estimates are based on 10^6 bootstraps of the data. For comparison, divergence data in Moyle and Nakazato (2010) are shown on the right, including our bootstrap analysis based on single-locus estimates given in Moyle and Nakazato's (2010) table S3 (unweighted K_s over six loci). n.a., not available.

large long-term effective population size in all other Solanaceous genera (Richman et al. 1996; Paape et al. 2008). Allelic diversity at the *S-RNase* locus is high in the wild tomato species *S. chilense* (Ilgic et al. 2007) and *S. peruvianum* (Miller and Kostyun 2011). Clearly, this diversity implies a large N_e for the common ancestor of all extant tomato species and the (older) common ancestor of the tomato clade and the nontomato outgroup *S. lycopersicoides*, the relevant historical entities for confounding ancestral polymorphism and interspecific divergence in this context.

The QTL data of Moyle and Nakazato (2010) are based on three interspecific comparisons, each using large numbers of introgression lines between the cultivated tomato, *S. lycopersicum* (SL), and each of the three wild species *S. pennellii* (SP), *S. habrochaites* (SH), and *S. lycopersicoides* (Slyc). Recent phylogenetic analyses have placed SP and SH as sister species to the remainder of tomatoes, which implies that both SP and SH share a single divergence time with the cultivated tomato, as their common ancestral species was basal to the entire tomato clade (Rodriguez et al. 2009). Although their study largely neglected the potential impact of intraspecific polymorphism on phylogeny reconstruction, Rodriguez et al. (2009) found extensive discordance among single-locus genealogies, even for alleles sampled from single heterozygous plants. Such patterns are consistent with incomplete lineage sorting and precisely what one would expect with large N_e and relatively recent divergence times. We thus chose not to unduly rely on the accuracy of the SP–SH sister relationship and evaluated the temporal accumulation of incompatibilities in two different ways: 1) based on three independently estimated K_s values (as done by Moyle and Nakazato 2010) and 2) by averaging the independent SL–SP and SL–SH divergence estimates to represent their putative common divergence time.

We reanalyzed molecular divergence between the focal *Solanum* species based on publicly available sequence data for 21 nuclear loci each with >31 synonymous sites ([supplementary table S1, Supplementary Material](#) online). In addition to calculating divergence between Slyc and SL, we also estimated this phylogenetically “wide” divergence based on sequences pooled across several tomato species versus the single Slyc sequence; the rationale being that the latter, nontomato outgroup species must share the same divergence time with all members of the tomato clade (Peralta et al. 2008; Rodriguez et al. 2009). Furthermore, given the limited number of synonymous-site differences across loci,

using many sequence comparisons to estimate the single underlying divergence time mitigates against biases due to lineage sorting and the stochastic nature of the substitution process (Gillespie 1991).

Our K_s estimates for the two comparisons SL–SP and SL–SH are similar to those of Moyle and Nakazato (2010), but our estimate for SL–Slyc is somewhat lower than their estimate of 0.0663 and almost identical to the estimated “tomato clade”–Slyc divergence (0.0541; [table 1](#)). Although the phylogeny-derived expectation of an equal divergence time for SL–SP and SL–SH is not strictly met by our K_s estimates, confidence intervals around the two K_s point estimates broadly overlap ([table 1](#)). Moreover, divergence estimates using the much larger number of silent sites exhibit only minute differences between the two species pairs ([supplementary table S1, Supplementary Material](#) online), consistent with a common divergence time.

Dobzhansky–Muller incompatibilities should increase with the square of interspecific substitutions or even faster depending on the complexity of the underlying deleterious gene interactions (Orr 1995; Orr and Turelli 2001). Hence, based on our K_s estimates and a wide range of ancestral polymorphism, we evaluated under which conditions the accumulation of seed sterility QTL and those underlying the other traits studied by Moyle and Nakazato (2010) is better approximated by a quadratic ($y = ax^2$) than by a linear model ($y = ax$). As illustrated in [figure 1](#) for the seed sterility QTL data, the impact of using separate K_s values for the SL–SP and SL–SH species pairs is minor, but the effects of assuming different levels of ancestral polymorphism and using alternate K_s estimates are profound. Specifically, the snowball effect has considerably more support under our divergence estimates, mainly as a consequence of Moyle and Nakazato's (2010) higher K_s estimate for the older SL–Slyc split ([table 1](#)). As the number of incompatibility QTL remains unchanged, our lower SL–Slyc divergence estimate implies a faster accumulation of incompatibilities and hence increased support for the quadratic model. In contrast, support for the quadratic model under Moyle and Nakazato's (2010) divergence estimates is only 59% even when no ancestral polymorphism is assumed and quickly decreases with increasing levels of ancestral polymorphism ([fig. 1](#)).

Just as ancestral polymorphism ought to have contributed to present-day sequence divergence between closely related species, polymorphism for incompatibility factors within species merits some attention. For outcrossing

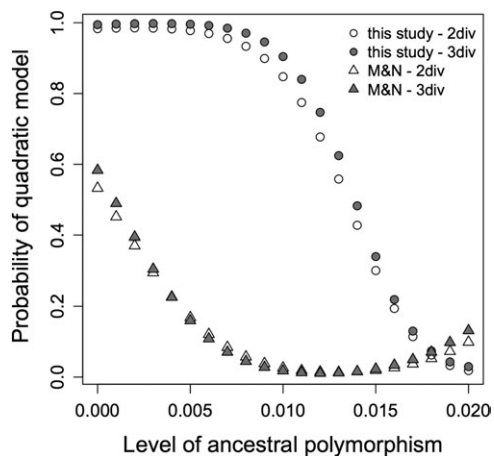


Fig. 1. Impact of ancestral polymorphism and divergence estimates on the probability that the quadratic model provides a better fit to the seed sterility QTL data of Moyle and Nakazato (2010) than a linear model. Probabilities were obtained by calculating Akaike weights (Burnham and Anderson 2002) from each model, using both the K_s values obtained by Moyle and Nakazato (2010) (“M&N”; triangles) and those based on our more extensive reanalysis (“this study”; circles). Open symbols track results when the average of the SL–SP and SL–SH K_s estimates is used as a proxy for the more recent divergence time (“2div”), and gray symbols plot the results when divergence estimates are kept separate between these two species pairs (“3div”). The gray triangle at $\pi_s = 0$ is equivalent to the Akaike Information Criterion values in Table 2 (model comparison 2) of Moyle and Nakazato (2010).

species, Dobzhansky–Muller incompatibilities are not expected to establish without geographic isolation, or they would be opposed by selection (the model appropriately posits divergence in allopatry; Coyne and Orr 2004).

However, polymorphism for incompatibility factors has been documented in geographically widespread plant species, foremost in the genus *Mimulus* (e.g., Christie and Macnair 1987; Sweigart et al. 2007; Scopece et al. 2010). Intriguingly, the most detailed study using natural *Mimulus* populations concluded that hybrid seed inviability does not commonly arise within and between species; in contrast, there is abundant evidence for partial hybrid pollen sterility both in inter- and intraspecific crosses, with the majority revealing patterns of inheritance consistent with the Dobzhansky–Muller model (Martin and Willis 2010). Although classical extensive crossing studies within and between wild tomato species were not designed to reveal the genetic underpinnings of post-zygotic barriers, they have demonstrated full intraspecific compatibility (assessed through hybrid seed failure, e.g., Rick 1986) at geographical scales comparable to studies of within-species nucleotide polymorphism. To the extent that these patterns in extant tomatoes and *Mimulus* are applicable to the ancestral tomato species, our formal assumption of zero whole-genome equivalent QTL at the time of species divergence introduces little bias for the focal trait, seed sterility; any such biases, however, would yield spurious support for the linear model.

Regardless of assumed levels of ancestral polymorphism, QTL for traits other than seed sterility appear to accumulate linearly or slower with time since divergence (supplementary fig. S1, Supplementary Material online). In stark contrast, interpreting the accumulation of seed sterility QTL is contingent on levels of ancestral polymorphism: Values lower than ~ 0.014 are needed to favor the quadratic over the linear model (figs. 1 and 2). Given the molecular data demonstrating high π_s in wild tomato species and the plausibility of high

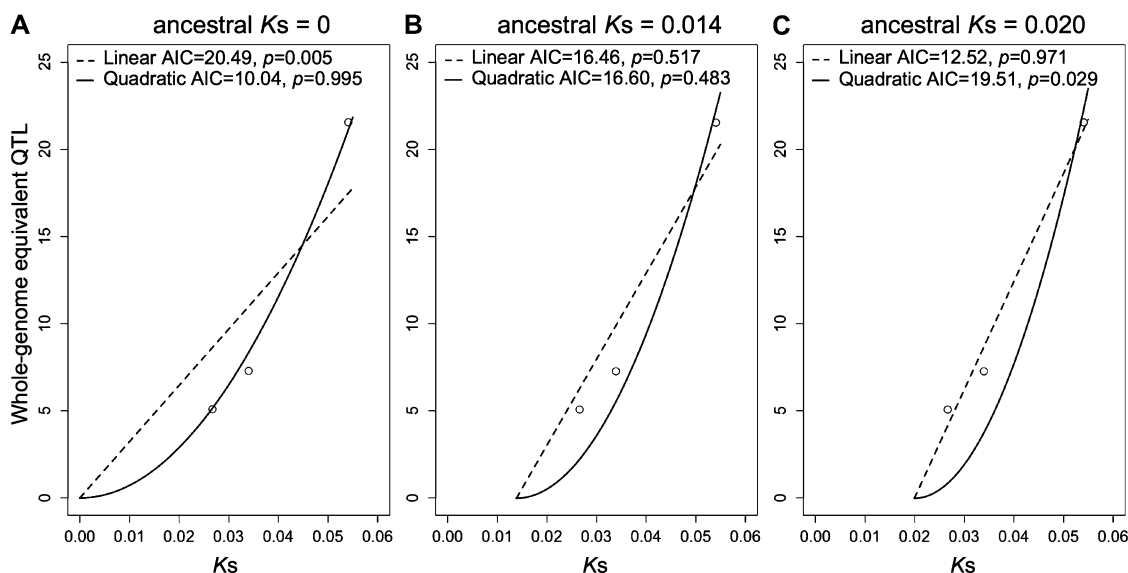


Fig. 2. Effect of ancestral polymorphism on the relative fit of linear (stippled lines) versus quadratic models (solid lines) for the accumulation of seed sterility QTL with evolutionary divergence in *Solanum*. Seed sterility QTL data from Moyle and Nakazato (2010) are plotted as open circles, their position along the x axis reflecting our K_s estimates. We show three scenarios representing different values of ancestral polymorphism at synonymous sites: (A) $K_s [\pi_s] = 0$, (B) $K_s = 0.014$, and (C) $K_s = 0.020$. From the range of ancestral π_s analyzed and shown in figure 1, these three scenarios were chosen simply to illustrate conditions under which the quadratic model provides a much better fit to the data (A), both models have about equal probability (B), and where the linear model provides a much better fit to the data (C). AIC, Akaike Information Criterion.

ancestral N_e (see above), we are inclined to conclude that currently available phenotypic and molecular data do not justify strong claims regarding the snowball effect in *Solanum*. In contrast, assuming reasonable levels of ancestral polymorphism still yields strong support for the snowball effect in *Drosophila* (supplementary fig. S2, Supplementary Material online), given the number of incompatibilities inferred by Matute et al. (2010). In addition to highlighting the impact of ancestral polymorphism on drawing divergent conclusions from identical raw interspecific divergence data in *Solanum*, our findings also illustrate how volatile inferences on the snowball effect are in the face of divergence estimates differing by only ~20% (0.0541 vs. 0.0663). We expect that more definitive genome-wide estimates of synonymous-site divergence between the focal *Solanum* species will help to resolve this latter issue in the near future.

Supplementary Material

Supplementary table S1, figures S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We are grateful to Sabine Güsewell for advice on statistical issues and the implementation of R scripts and thank three anonymous referees for helpful comments on a previous version of this manuscript. This work was supported by the Swiss National Science Foundation (grant 31003A_130702 to T.S.) and the Institute of Integrative Biology at ETH Zurich.

References

- Arunyawat U, Stephan W, Städler T. 2007. Using multilocus sequence data to assess population structure, natural selection and linkage disequilibrium in wild tomatoes. *Mol Biol Evol.* 24:2310–2322.
- Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York: Springer.
- Christie P, Macnair MR. 1987. The distribution of postmating reproductive isolating genes in populations of the yellow monkey flower, *Mimulus guttatus*. *Evolution* 41:571–578.
- Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.
- Gillespie JH. 1991. The causes of molecular evolution. Oxford: Oxford University Press.
- Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B. 2010. Species selection maintains self-incompatibility. *Science* 330:493–495.
- Hey J. 2010. Isolation with migration models for more than two populations. *Mol Biol Evol.* 27:905–920.
- Igic B, Bohs L, Kohn JR. 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proc Natl Acad Sci U S A.* 103:1359–1363.
- Igic B, Smith WA, Robertson KA, Schaal BA, Kohn JR. 2007. Studies of self-incompatibility in wild tomatoes: I. S-allele diversity in *Solanum chilense* Dun. (Solanaceae). *Heredity* 99:553–561.
- Li W-H. 1977. Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics* 85:331–337.
- Martin NH, Willis JH. 2010. Geographical variation in postzygotic isolation and its genetic basis within and between two *Mimulus* species. *Phil Trans R Soc Lond B.* 365:2469–2478.
- Matute DR, Butler IA, Turissini A, Coyne JA. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* 329:1518–1521.
- Miller JS, Kostyun JL. 2011. Functional gametophytic self-incompatibility in a peripheral population of *Solanum peruvianum* (Solanaceae). *Heredity* 107:30–39.
- Moyle LC, Nakazato T. 2010. Hybrid incompatibility “snowballs” between *Solanum* species. *Science* 329:1521–1523.
- Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics* 139:1805–1813.
- Orr HA, Turelli M. 2001. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution* 55:1085–1094.
- Paape T, Igic B, Smith SD, Olmstead R, Bohs L, Kohn JR. 2008. A 15-myrr-old genetic bottleneck. *Mol Biol Evol.* 25:655–663.
- Peralta IE, Spooner DM, Knapp S. 2008. Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Syst Bot Monogr.* 84:1–186.
- Peterson GI, Masel J. 2009. Quantitative prediction of molecular clock and Ka/Ks at short timescales. *Mol Biol Evol.* 26:2595–2603.
- Richman AD, Uyenoyama MK, Kohn JR. 1996. Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science* 273:1212–1216.
- Rick CM. 1986. Reproductive isolation in the *Lycopersicon peruvianum* complex. In: D’Arcy WG, editor. Solanaceae—biology and systematics. New York: Columbia University Press. p. 477–495.
- Rodriguez F, Wu F, Ané C, Tanksley S, Spooner DM. 2009. Do potatoes and tomatoes have a single evolutionary history, and what proportion of the genome supports this history? *BMC Evol Biol.* 9:191.
- Roselius K, Stephan W, Städler T. 2005. The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* 171:753–763.
- Scopece G, Lexer C, Widmer A, Cozzolino S. 2010. Polymorphism of postmating reproductive isolation within plant species. *Taxon* 59:1367–1374.
- Städler T, Arunyawat U, Stephan W. 2008. Population genetics of speciation in two closely related wild tomatoes (*Solanum* section *Lycopersicon*). *Genetics* 178:339–350.
- Städler T, Roselius K, Stephan W. 2005. Genealogical footprints of speciation processes in wild tomatoes: demography and evidence for historical gene flow. *Evolution* 59:1268–1279.
- Sweigart AL, Mason AR, Willis JH. 2007. Natural variation for a hybrid incompatibility between two species of *Mimulus*. *Evolution* 61:141–151.
- Tellier A, Fischer I, Merino C, Xia H, Camus-Kulandaivelu L, Städler T, Stephan W. 2011. Fitness effects of derived deleterious mutations in four closely related wild tomato species with spatial structure. *Heredity* 107:189–199.
- Wakeley J. 2009. Coalescent theory. Greenwood Village (CO): Roberts & Company.