Evolutionary Genetics of an *S*-Like Polymorphism in Papaveraceae with Putative Function in Self-Incompatibility

Timothy Paape[1]*, Takashi Miyake[2], Naoki Takebayashi[2,3], Diana Wolf[2,3], Joshua R. Kohn[4]

1 College of Biological Sciences, University of Minnesota, St. Paul, Minnesota, United States of America, 2 Institute of Arctic Biology, University of Alaska, Fairbanks, Alaska, United States of America, 3 Department of Biology and Wildlife, University of Alaska, Fairbanks, Alaska, United States of America, 4 Department of Biology, University of California San Diego, La Jolla, California, United States of America

Abstract

Background: *Papaver rhoeas* possesses a gametophytic self-incompatibility (SI) system not homologous to any other SI mechanism characterized at the molecular level. Four previously published full length stigmatic *S*-alleles from the genus *Papaver* exhibited remarkable sequence divergence, but these studies failed to amplify additional *S*-alleles despite crossing evidence for more than 60 *S*-alleles in *Papaver rhoeas* alone.

Methodology/Principal Findings: Using RT-PCR we identified 87 unique putative stigmatic *S*-allele sequences from the Papaveraceae *Argemone munita*, *Papaver mcconnellii*, *P. nudicaule*, *Platystemon californicus* and *Romneya coulteri*. Hand pollinations among two full-sib families of both *A. munita* and *P. californicus* indicate a strong correlation between the putative *S*-genotype and observed incompatibility phenotype. However, we also found more than two *S*-like sequences in some individuals of *A. munita* and *P. californicus*, with two products co-segregating in both full-sib families of *P. californicus*. Pairwise sequence divergence estimates within and among taxa show *Papaver* stigmatic *S*-alleles to be the most variable with lower divergence among putative *S*-alleles from other Papaveraceae. Genealogical analysis indicates little shared ancestral polymorphism among *S*-like sequences from different genera. Lack of shared ancestral polymorphism could be due to long divergence times among genera studied, reduced levels of balancing selection if some or all *S*-like sequences do not function in incompatibility, population bottlenecks, or different levels of recombination among taxa. Preliminary estimates of positive selection find many sites under selective constraint with a few undergoing positive selection, suggesting that self-recognition may depend on amino acid substitutions at only a few sites.

Conclusions/Significance: Because of the strong correlation between genotype and SI phenotype, sequences reported here represent either functional stylar *S*-alleles, tightly linked paralogs of the S-locus or a combination of both. The considerable complexity revealed in this study shows we have much to learn about the evolutionary dynamics of self-incompatibility systems.

Introduction

Angiosperms possess various genetic mechanisms to reduce inbreeding by rejecting self-pollen. Single locus gametophytic self-incompatibility (GSI) occurs in several plant families [1,2] and is defined by rejection of haploid pollen in the stigma or style if the pollen carries a self-incompatibility allele (*S*-allele) that matches either allele of the ovule parent. Crossing studies [3] of self-incompatible *Papaver rhoeas* estimated that this species maintains 66 *S*-alleles at the gametophytic *S*-locus [4]. Molecular studies led to the cloning of two stigmatic *S*-alleles [5], but despite considerable efforts to clone additional haplotypes, Kurup *et al.* [6] were only able to sequence one additional *S*-allele from *P. rhoeas* (*S8*) and one

from the congener *P. nudicaule* (*Sn1*). It is thought that the very high level of divergence among *Papaver* *S*-proteins confounded the isolation of more *S*-alleles by standard PCR and hybridization techniques [6]. Sequence analysis of these alleles and studies characterizing the molecular mechanism of pollen tube inhibition revealed that this system bears no homology to the *S*-RNase-based gametophytic self-incompatibility system found in the Solanaceae, Plantaginaceae and Rosaceae, or the sporophytic SI system (SSI) found in Brassicaceae where the principle stylar product of the *S*-locus is an *S*-receptor kinase (*SRK*; reviewed in [7,8]).

The *S*-locus products secreted in the pistil of *Papaver* are small proteins (~15 kDa) that bind to corresponding pollen *S*-receptors when haplotypes match, initiating receptor mediated programmed

cell death [9–12]. The *Papaver* stigmatic *S*-proteins possess four conserved cysteine residues and the predicted secondary structure is comprised of ß-strand motifs linked by seven hydrophilic loops [13,14]. Other types of molecular recognition proteins [15–17] have distinct hypervariable regions interrupted by conserved regions contrasting with *Papaver* stigmatic *S*-proteins where variability is at the 5′ end (∼30 amino acids) and at scattered residues throughout the sequence. In *Papaver* stigmatic *S*-proteins display 40–50% amino acid divergence between pairs of alleles, as do the recently-discovered pollen *S*-proteins they interact with [12]. An early scan of the *Arabidopsis* genome revealed as many as 100 '*S*-protein homologues' (SPHs; [18]) and another SPH was described in tomato [19]. SPHs are the closest known homologs of the *Papaver* stigmatic *S*-protens but no function has yet been ascribed to any SPH.

The *S-RNase* and *SRK* systems both exhibit extensive ancestral polymorphism shared across genera [16,20,21]. This reflects balancing selection preserving variation at the self-incompatibility (*S-*) locus for millions of years [22,23] and makes it possible to infer demographic events that occurred before the origin of extant species [21,24–26]. The extent of shared ancestral *S*-locus polymorphism in the Papaveraceae is unknown because, while many other Papaveraceae are self-incompatible [27–30], *S*-locus sequence polymorphism has not been investigated outside of *Papaver* [6].

Here we attempt to isolate homologous *S*-alleles from SI species of Papaveraceae, focusing primarily on wild relatives of *Papaver rhoeas*. Three species native to California are investigated: *Argemone munita*, *Platystemon californicus*, and *Romneya coulteri*. Both *A. munita* and *R. coulteri* are tetraploid and while the former typically grows as a desert annual, the latter is a perennial and often grows clonally. *Platystemon californicus* is a diploid summer annual and the only member of its genus. Self-incompatibility reactions of all three native Californian species have been examined using aniline blue staining and other microscopy techniques [31]. In all cases self pollen tubes arrest immediately upon contact with living female tissue, as is found in *Papaver*. We also look for *S*-like polymorphism in two Alaskan species: *Papaver mcconnellii*, a mostly tetraploid perennial that grows on rocky scree in the Yukon Territory [32] and *P. nudicaule*, an annual or short-lived perennial that may be diploid or tetraploid. Our goals in this study were to: a) amplify putative *S*-alleles from these species based on known *Papaver S*-allele sequences, b) determine whether putative *S*-haplotypes predict the SI phenotype in greenhouse crosses, c) assess genealogical characteristics and patterns of sequence variation in putative *S*-alleles within and among species.

Results

For most taxa studied, stigmatic sequences of putative *S*-alleles were amplified from stigmatic tissue using RT-PCR althought some sequences from *Papaver* spp. were amplified from genomic DNA (see methods). Sequence data from have been deposited with the EMBL/GenBank Data Libraries under accession nos. (GQ351602–GQ351691).

The sequences amplified were highly polymorphic as expected for putative *S*-alleles and showed the nearest match in NCBI nucleotide tBLASTx queries to the published *S*-alleles of *Papaver*. Thirty-seven complete or partial sequences were amplified from *A. munita*. We amplified two sequences from 12 *A. munita* individuals, only one sequence from 10, while in one plant three sequences were found. A total of 31 partial *P. californicus* sequences were obtained using 3′ RACE and nested PCR and eleven of these sequences were completed using 5′ RACE (Figure 1). For *P.*

californicus, two products were found for 11 individuals, only one product was found for 10 individuals and three or four products were found for 4 individuals. Two sequences were also found for all 12 *Romneya coulteri* sampled for a total of six unique sequences. PCR from stigmatic tissue resulted in six new full-length putative *S*-alleles from *P. mcconnellii* and one from *P. nudicaule*. For *Papaver* species, only one allele was detected for each of the seven individuals tested.

Hand pollinations —*Argemone munita*

Two *A. munita* full-sib families were genotyped ('8-1' and '25-4') using allele specific primers for the 5′ ends of putative *S*-allele sequences. Family 8-1 was generated by crossing plant 8 (with alleles *S8a* and *S8b*) with plant 1 (with alleles *S1c* and *S1d*), while family 25-4 was generated by crossing plant 25 (*S25a* and *S25b*) with 4 (*S4a* and *S4d*). Here, we denote putative *S*-alleles with plant individual ID (digits; e.g., 4) and a single letter (e.g., 'a', 'c' or 'd') indicating different sequences amplified from the same individual. Sequences segregated in an allelic fashion in both F1 families (Table S1). When crossing full-sibs, we expect that crosses among plants with only one shared allele (predicted semi-compatible) will set fruit, whereas crosses between matching genotypes (predicted incompatible) will not. As expected, 100% of the 64 pollinations among individuals possessing only one shared allele produced fruit, while none of 15 matching-genotype crosses from family 8-1 set fruit (Table 1). However, in family 25-4, some matching-genotype crosses set fruit while others did not (Table 1). Unexpected fruit set only occurred in crosses among plants that had the sequence *S25b*. We recorded 59% fruit set in 34 matching-genotype pollinations involving putative allele *S25b* and zero fruit set in 43 pollinations not involving *S25b* (Table 1). This could suggest that *S25b* is a weak or non-functional *S*-allele or that the *S25b* sequence is not linked to the functional *S*-locus. Overall, this set of crossing data supports that at least seven out of eight putative *S*-alleles that segregated as heterozygous products (*S25a, S4a, S4d, S8a, S8b, S1c, S1d*) are at least linked to functional *S*-alleles.

Autogamous fruit set was rare among greenhouse grown *A. munita* plants despite large deposits of self-pollen visible on stigmas. However, some autogamous seed set occurred on plant 25, the parent that carried the *S25b* allele, as well as in 3 of 21 of its progeny favoring the hypothesis that *S25b* is a weak incompatibility allele. Occasional fruit set was also observed among un-genotyped bagged and hand self-pollinated individuals of *A. munita* (7 of 61 plants) indicating either additional variation in the strength of SI or that allele *S25b* was present among these plants. Outcross hand pollination from non-sib donors nearly always set fruit (136 of 141 pollinations). None of the F2 offspring from crosses among siblings germinated, perhaps due to inbreeding depression in addition to the already low rates of germination common in Papaveraceae. It was therefore impossible to test for expected segregation of parental sequences among F2 offspring.

Hand pollinations – *Platystemon californicu*:

In contrast to *A. munita*, fruit set in *P. californicus* was never observed following hand or autogamous self-pollination but occasional fruit set did occur after putatively incompatible crosses. Prior to genotyping the progeny from two full sib families (1–2 and 1–3), we identified two stigma-S sequences from each parent: *S1a* and *S1b* from parent 1, *S2d* and *S2e* from parent 2, and *S3a* and *S3b* from parent 3. However, in both crosses, we later found that some of the putative *S*-sequences failed to segregate independently, but rather co-segregated as if linked. In family 1-3, sequences *S3a* and *S3b* (hereafter *S3a,b* where *a* and *b* are unique PCR products) were both found in 12 of 22 individuals and never found

```
                  10         20         30         40         50         60         70         80
         ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|
PrsS8     ------MKLL  YAILFLSFLT  LASSRFL--P  VIEVRIMNKR  GNGHSIGIHC  RSKDDDLGYH  RISDGQQVHF  SFRENFFHTT
PrsS1     ------.NIF  .V.VL...FL  SK..G.F--.  ........R.  ...R.....  .....N.QNQ  TVTS.HDMS.  ....D....
PnSn1     ------.T.   FVFTL.F..C  MSF.DGS-RS  RVQ..V..R.  ...KTVE..   Q.G.....N.  VVP...E.N.  ....S.LEN.
PrsS3     ------.I.   CV.VL.TLVS  MSN.IGF--T  R.Q....R.   ...K..T..   Q.E....T.   VV....EINW  .....ML...
PmccS2    ------.GIF  .VVVL.C.FL  ST.NGWF--.  D......R.   .S.K....   .....N.QNQ  TVIS.HDMS.  ..K.D.W...
PmccS6    ------.T.   FVVSL.L..C  MSF.DGS-RS  RVQ..V..R.  ...KTVE..   Q.G.....N.  VV....E.N.  ....S..EN.
PmccS8    ------.T.   FVVSL.F..C  MSF.NAV---  H.Q..V..R.  ...KTVE..   Q.G.....N.  VVP...E...  ..G.S..EN.
PmccS4b   ------.T.   FLVSL.F..C  MSF.NAA---  H.Q..V..SG  ...KTVE..   Q.G.....N.  VVP...E.NW  .GVS..EN.
PmccS5a   ------.TF   FVVSLVF..C  .CI.FNG--.  H.Q..V..R.  ...KTVE..   Q.G.....N.  VV....E.N.  .G.S..EN.
PmccS7    ------.RTF  .V..L.LV.C  FD..NGR---  R.HI....RI  ...K.MP..   Q.......NQ  IV....D.QW  ....SLLLN.
AmS1c     MNIKNTL.PY  D..C.AIP.L  .DPVQCS-LT  P.HAS....R.  H..K..EV..  Q.V.N.DNQ  VVA..NE.KW  ...S..EDA
AmS1e     MNIKSTF.PY  NV.W.TIL.L  .NQVQ---LT  P.HA.....R.  ...K..EV..  Q.V.N.NQ   VVP..NE.KW  T...SLLEN.
AmS25A    MNIQNTLRPY  I..W.AIL.Q  .DQVQSS-LT  P.QA.....R.  ...K..EV..  K.L.N.NQ   VVA.DNELRW  T.K.S.WQN.
AmS25B    MYTKNTL.PC  NV.W.AI..L  .DPVQSH-LR  P.H......R.  ...K..EV..  K.V.N.DNQ  IVA..NEA.W  .....SLLEN.
AmS8E     MTIKNALTPY  N..W.VIL.L  .VPVQCS-LT  P.QA.....R.  ...K.VE..   K.L.N.DNQ  VLL..NELKW  K.K.SIWEN.
AmS8F     MNRENTL.PY  N.LW.AIL.L  HHPVESS-LT  P.QA.....R.  ...K..EV..  Q.V.N.DNQ  VVA..NELKW  T...T.WEN.
AmS18F    MNIKNTL.PY  N..C.AIL.L  .DQVQSN-LT  P.N......R.  ...K..EV..  Q.V.N.DHQ  VVA..NE.KW  T...G..EN.
AmS17I    MNFKNTL.PY  K..WLAIL.P  .DPVQSS-LT  P.H......R.  ...K..VE..  ..A.N.DNQ  VVL..NEIKW  T.K.S.WEN.
AmS17C    MNIKNTF.PY  N..R.AIL--  .DPVQSS-LT  P.H.....RK  .D.K.VE..   Q.G.....NQ  VVA..NELKW  N...S.YEN.
AmS4a     MNTKNTL.PY  NV.C.AII.L  .DQVQSN-LT  P.H......R.  ...K..EV..  Q.V.N.DNQ  VVP..NE.KW  T...IWGN.
AmS4b     MYTKNT---Y  N.TC.AIL.L  .DQVQSN-LT  P.HA.....R.  ...KN.EV..  Q.G.N.DNQ  IVA..NE.RW  T.K.SL.ED.
AmS20H    MNTRNT--PY  N....AIL.L  IDPVHPS-LT  P.QA.....R.  ...K.VEV..  K.L.N.DNQ  VVA..NELKW  T.K.S.WEN.
Pcal1_3g  MNKKKNLR.N  F.KW.VI..L  .DQVQSS-LT  P.YA.....R.  A..K..EV..  ....NQ     VVL..NELKW  T.K.H.YQD.
Pcal1_3b  MNTKKTL.VY  FT...SILFL  .HQVHSS-LR  L.HA.....R.  N..K..EV..  ..F...NNQ  LVL..NELAW  E.K.S.NEE.
Pcal1_3e  IKMKKML.IY  IT...AII.L  IDQVHSN-LR  L.HA.T..R.  D..K..EV..  ....NNQ    VVL..NELAW  K...R..QD.
Pcal1_1b  TNMRNTLRS.  N..W.AILIL  FEHGQAS-YV  P.H......R.  ...K.G.N..  K.G.N.NQ   VVL..DEQKW  K.K.S.WEN.
Pcal5_1G  INMKNTL.SF  N..W.AVLIL  .EHGQAN-YV  P.H......R.  ...K..EV..  K.G.....ENQ VVL..DEQKW  T.K.S.WEN.
Pcal5_1a  RNMKKTL.SY  N..W.AI.IL  .EHGQSS-LT  P.Y......R.  ...KN.E...  K.A.N.NQ   VVL..GEQKW  T.K...WE.
Pcal6_11b MNTKKTL.VY  FT...SILFL  .HQVHSS-LR  L.HA.....R.  N..K..EV..  ..F...NNQ  LVL..NELAW  E.K.G.NEE.
Pcal9_2a  --MNTYL.PY  NV.W.TIL.L  .EQGQSSILV  PTH......R.  ...K..AV..  L.V.....QNQ VVL..DEQSW  T...HIYQD.
Pcal9_3f  MNMKKNLNIY  ITL..AL..L  FDQVHSG-VR  ..HA.....R.  N..E..EV..  W..E...NNQ  VVL..NEIAW  K...SVLQN.
Pcal6_11d LKMKKTL.VY  AI...VIL.L  VDQTHCS-LR  L.HA.....R.  ...K..EV..  W.A...NNQ   VVA..HE.AW  K...HI.QD.
Pcal34L   --MKNSLR.Y  F..WLAI..L  IEQGQSS-LT  P.YA.....R.  .S.K..EV..  .......NQ  VLL..NEQNW  T...HL.EN.


                  90         100        110        120        130        140
         ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....|....|  ....
                                       ▼
PrsS8     TFNCDTFWD-  SRRHFNFDSY  RAQRDDHGRC  TTFCLWKTTD  FGLYGYDQFH  FYWQLYYLAK  K---
PrsS1     H.Y..LQ..-  KETK.G.Y.   Q.K...D...  SSQ....IM.  D....F...   QI..I.H.I.  .E--
PnSn1     R.W..VQ.S-  NEVKYH..A.  WSD..RL...  LSQ....TMME D.......K   QV.E.VH..V  .E--
PrsS3     R.Y.YLQ..R  KGK.....A.  SGK....G...  S......ILE  G..W....Q   QI..AFS.PG  MG--
PmccS2    H.Y..LQ..-  KETK.G.Y.   Q.K...D...  NSQ....VL.  D....F...   QI..T.H.TN  .ER-
PmccS6    R.W..VQ.S-  .EIKYH..A.  WSK..NL...  LSQ....TMRE D.......K   QI..VH..V   .E--
PmccS8    R.W..VQ.S-  .EIK.H.F.A. WSD..RL...  D.Q....TMNE D.......K   QV.E.VH..V  .G--
PmccS4b   R.W..VK.S-  .EVKYH..A.  WTK...L...  SSQ....TMRE D......N    QL...VH..V  .A--
PmccS5a   R.W..VK.S-  .EVKYH..A.  WTK..AL...  SSQ....TMKE D.......K   QI.E.VH..V  .G--
PmccS7    H.Y..LQ.N-  GAANVH..AF  WSK..HL...  .SQ....IVQ  K..L.....Q  AK.EI....S  ATE-
AmS1c     R.Y..LQ.N-  KTIK.H..A.  WSD..HT...  FS.....VSE  D..-----   ----------  ----
AmS1e     R.Y.Y.Q.N-  EAI..H..A.  WSK...S...  SS.....I.E  D..-----   ----------  ----
AmS25A    R.T..LK.N-  ETIN.H.NAF  FST...W..V  SSK....VQ.  D..-----   ----------  ----
AmS25B    R.Y..LK.N-  DTIK.H.NA.  WSE..GW...  SS.....I.E  D.F?.-----  ----------  ----
AmS8E     R.S..LK.N-  ETYS.H..A.  WSV..NM...  F.K....ILE  D..F.-----  ----------  ----
AmS8F     R.Y..LK.N-  ETTK.H.NA.  WSD..HW..I  SSK....VQE  D..------  ----------  ----
AmS18F    R.Y..LL.N-  MSIN.H..A.  WSD...S...  SSQ....ILE  D........N  KI..IL...V  .DK-
AmS17I    R.S..LQ.N-  KEIS.H..A.  WSV..NG...  F.I....IME  D..F...E.N  KV..IL...V  EDK-
AmS17C    L.Y..LS.N-  ETIK.H..A.  WSE..NG...  F.I....I.E  D..F.N.E.N  KI..IL...V  .DK-
AmS4a     R.F.YLQ.N-  .TFN.H..A.  MSD...S...  F......ILE  .........N  KI...F...V  .DK-
AmS4b     R.S..LK.N-  ETIK.H.YA.  KSE...W..V  NS.....VLE  D........N  .I..IF..TV  .DK-
AmS20H    R.T..LK.N-  ETI..H.NAF  WSK...W..V  S......VQE  D..-----   ----------  ----
Pcal1_3g  R.T...S.S-  KTIK.H..A.  WSE..ND...  F.I.Y.EVIR  GS...-----  ----------  ----
Pcal1_3b  Q.Y..LK.N-  DTIK.H..A.  LSR...D...  FKI.YYQVLE  ......E.T   KI.-------  ----
Pcal1_3e  H.Y..LK.N-  ETYKYH..A.  LSK...Y...  F.I.YYQVL.  ......E.N   KI.L.I.V.-  ----
Pcal1_1b  L.Y.NLK.N-  EEVFSP..A.  DSD..NF...  FSI.Y.LV.E  V......E.N  QI.-------  ----
Pcal5_1G  L.Y.NLK.N-  EEVFSP..AF  DSE..NI...  F.I.Y.LV.E  I......E.S  KV.-------  ----
Pcal5_1a  L.F.YLK.N-  EEVY.K..V.  DSD...G...  F.I.Y.TV.E  V......E.T  NI.-------  ----
Pcal6_11b Q.Y..LK.N-  DTIK.H..A.  LSR...D...  FKI.YYQVLE  .......EGN  KI.-------  ----
Pcal9_2a  R.I..LK.S-  EAIT.H..A.  WSY...Y...  FSK.Y.QV.E  NA.....E.N  QI.-------  ----
Pcal9_3f  R.Y..LK.N-  ETFTYH..A.  LSK...Y...  FSK.YYQVLE  .......E.N  QI.-------  ----
Pcal6_11d R.Y..LK..-  DTTK.H..G.  WSD...Y...  F.K.Y.QV.E  KC.....EGN  NI.-------  ----
Pcal34L   R.A..LK..-  ETIM.H..A.  WSN..NY...  F.I.Y.QV.E  NA.....E.N  QIREIVH..R  .DSK
```

Figure 1. Amino acid alignment of all sequences containing 5′ ends. All ten *Papaver* spp. sequences shown are full length possessing both 5′ and 3′ ends. Dots are invariable amino acids, gray indicates 40% shared identity threshold. Black arrow indicates the amino acid (site 107) used in the site directed mutagenesis experiment ([14]; see text) which is predicted to be under positive selection by both PAML [73] and OmegaMap [34] analyses.
doi:10.1371/journal.pone.0023635.g001

separately, ruling out that they are in fact separate alleles of a heterozygous *S*-locus as had been assumed when the cross was performed. A third product, *S3c*, was found but only observed in the absence of *S3a,b*. No homologous product linked to *S3c* was detected. Sequences *S1a* and *S1b* from the other parent in this cross were never found together in F1 progeny, characteristic of alleles segregating at a heterozygous *S*-locus. Family 1-2 also possessed a similar co-segregating polymorphism among F1 progeny. The PCR products *S2d and S2e* (hereafter *S2d,e*) from the male parent of this family were found together in 13 of 22 individuals (Table S1), while no sequence from plant 2 was found in the remaining 9 offspring. We apparently failed to amplify the alternative allele in this case.

Hand pollinations within the two families resulted in 93% fruit set when parents had one non-matching allele, while only 12% of crosses among plants with matching putative *S*-locus genotypes set fruit (Table 2). In *P. californicus*, a total of 16 fruits were set after putatively incompatible crosses. Ten of these fruits occurred in crosses among sibs from Family 1-3 where only one parental allele (*S1b*) was amplified and no sequence contributed by parent 3 was recovered (denoted S1b*×S1b* in Table 2). Fruit set among these offspring may have been due to failure to correctly genotype some individuals. If crosses among these ambiguous genotypes are excluded, only 6 of 108 pollinations among putatively incompatible individuals set fruit.

Fruit set following 22 hand self-pollinations in *P. californicus* was never observed (data not shown) nor was there any autogamous fruit set of bagged flowers despite copious autogamous self-pollen on stigmas. Germination of seeds from one cross among F1's with non-matching genotypes was sufficient to perform crosses among F2's. In Family 1-2, F2 crosses among offspring with the *S1a S1b* heterozygous genotype produced no fruit (N = 17, Table 2). In both *A. munita* and *P. californicus*, fruit set was significantly higher when non-matching rather than matching genotypes were crossed (Fisher's exact tests: *A. munita*, $\chi^2 = 93.02$ (1 *df*), $p<0.001$; *P. californicus* $\chi^2 = 113.74$ (1 *df*) $p<0.001$).

The finding of more than two *S*-like sequences from single individuals presents the possibility of a multi-locus system in which more than four incompatibility phenotypes might result from a cross. However, we always found incompatible pairs of plants in diallels among five full sibs indicating no more than four incompatibility phenotypes were present in each group of five sibs. Three full-sib 5×5 crossing arrays of *A. munita* all had at least one pair of individuals that were reciprocally incompatible (data not shown). An additional array possessed one pair of plants where three pollinations in one direction failed to produce fruit but the pair was not crossed reciprocally. In *P. californicus*, four arrays among full-sib family 1-3 and two among family 1-2 each possessed at least one pair of reciprocally incompatible plants. While this test cannot rule out that more than four mating types are present in a full sibship, we found no evidence of more than four mating types in any of the ten sets of five sibs tested, consistent with expectations of single-locus GSI.

Tissue specific RT-PCR

We investigated the expression of putative *S*-alleles in various tissues by allele specific amplification of cDNA (Figure 2). Overall, the expression of putative *S*-alleles was mostly localized to the stigma. The *A. munita*, sequences *S4a*, *S4d*, *S8a*, *S8b*, *S25a* and *S25d* showed no expression in leaves but appear to be highly expressed in stigmatic tissue (Figure 2). Other tissue types were not tested for this species. In *P. californicus*, we used an individual (genotype *S1b S2d,e*) from the crossing family 1-2 (Table 2). Alleles *S1b* and *S2d* showed high expression in the stigma as expected for a *S*-allele, with a faint band appearing in the leaf tissue for allele *S1b* (Figure 2*b*). Allele *S2e*, a potential paralog linked to *S2d*, appeared to show weak expression in unopened buds, stems, and leaves, as well as in the stigma. However, sequencing of this amplicon from non-stigmatic tissue types revealed that it was not *S2e* but a potential paralog not previously identified from the parental stigmas.

Table 1. Fruit set following hand pollinations among full sibs of *Argemone munita* with matching genotypes (predicted incompatible) or with only one matching allele (bold, predicted semi-compatible).

Predicted Incompatible				Predicted Semi-compatible	
Family 8-1		**Family 25-4**			
Genotype	Fruit Set	Genotype	Fruit Set	Genotypes	Fruit Set
S8a S1c×S8a S1c	0/7	*S25aS4a×S25aS4a*	0/32	*S8a **S1c**×S8b **S1c***	7/7
S8b S1d×S8b S1d	0/5	*S25bS4a×S25bS4a*	8/16	*S8a **S1d**×S8b **S1d***	7/7
S8b S1c×S8b S1c	0/3	*S25aS4d×S25aS4d*	0/11	*S4a **S25a**×S4d **S25a***	20/20
		S25bS4d×S25bS4d	12/18	***S4a** S25a×**S4a** S25b*	12/12
				*S4a **S25b**×S4d **S25b***	18/18
total	0/15		20/77		64/64

Parents of full sib Family 8-1 had genotypes *S8a S8b* and *S1c S1d*.
Parents of family 25-4 parental genotypes were *S25a S25b* and *S4a S4d*.
Predicted incompatible crosses failed to set fruit unless parents possessed allele *S25b* while all predicted semi-compatible crosses set fruit. Family names denote the plant numbers of the parents. For instance family 8-1 are offspring derived from crosses between plant 8 plant 1.
doi:10.1371/journal.pone.0023635.t001

Table 2. Fruit set following hand pollinations among full sibs of *Platystemon californicus* with matching genotypes (predicted incompatible crosses) or one non-matching allele (bold, predicted semi-compatible crosses).

Predicted Incompatible				Predicted Semi-compatible	
Family 1-3		**Family 1-2**			
Genotype	**Fruit Set**	**Genotype**	**Fruit Set**	**Genotypes**	**Fruit Set**
S1a S3a,b×S1a S3a,b	0/33	S1a S2d,e×S1a S2d,e	4/18	**S1**a S3a,b×**S1**a S3c	16/17
S1b S3a,b×S1b S3a,b	0/17	S1b S2d,e×S1b S2d,e	1/13	**S1**b S3a,b×**S1**b S3c	18/20
S1a S3c×S1a S3c	1/8	(F2) S1a S1b×S1a S1b	0/17	S1a **S3a,b**×S1b **S3a,b**	7/7
S1b*×S1b*	10/23			**S1**b S2d,e×**S1**b S2?	5/5
				S1a **S2d,e**×S1b **S2d,e**	10/11
total	11/81		5/48		56/60

Parent 1 was genotyped S1a S1b, parent 3 was S3a,b S3c where S3a,b appears to be a linked polymorphism and S3c is allelic to one or both S3a,b sequences (see text).
Parent 2 was genotyped S2d,e S2? where sequences d,e were always either inherited together in F1's or, when neither was present, no alternative product was found.
*indicates individuals where only one S-like sequence was recorded.
F2 genotype S1a S1b was derived from predicted semi-compatible crosses among full-sibs from family 1-2.
Family names as in Table 1.
doi:10.1371/journal.pone.0023635.t002

Phylogeny

Phylogenetic reconstruction of sequences from *Papaver*, *Argemone*, *Platystemon* and *Romneya* shows interesting features (Figure 3). The previously isolated *S*-alleles from *Papaver* exhibit no shared ancestral polymorphism with non-*Papaver* sequences. The newly identified *P. mcconnellii* and *P. nudicaule* alleles show some trans-
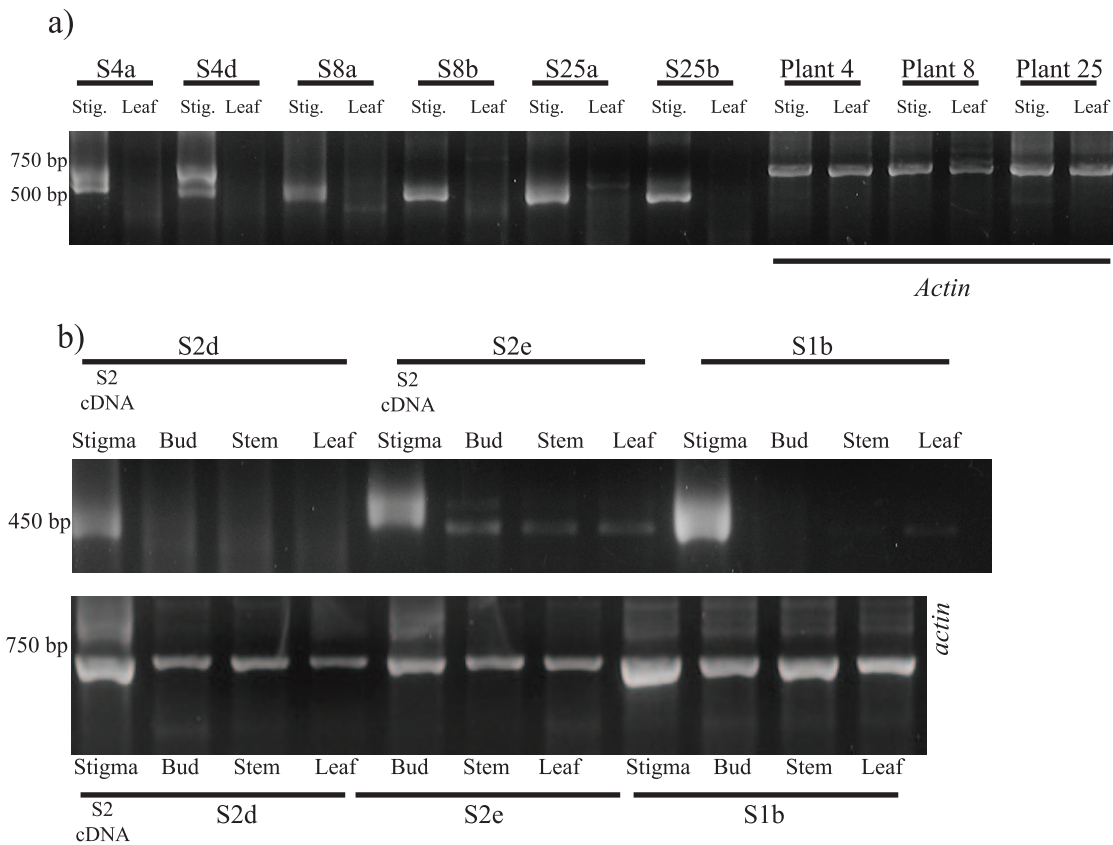

Figure 2. RT-PCR of putative *S*-locus genes from stigmatic, stem and leaf tissues using allele-specific forward primers and a universal amplification reverse primer. (a) Expression of *A. munita* sequences was detected in stigmatic but not leaf tissue. (b) Expression in *P. californicus* tissues is highest in stigmas as expected for *S*-alleles. Actin primers were used as positive controls for all plant tissue cDNA from both species. Putative alleles *S2d* and *S2e* appear to be linked. Sequencing of bands in non-stigmatic tissue using *S2e* primers showed that they are products not previously identified from stigmas. Actin primers were used only once on plant S2 stigmatic tissue, hence there is one less lane in the actin gel than in the allele specific gel.
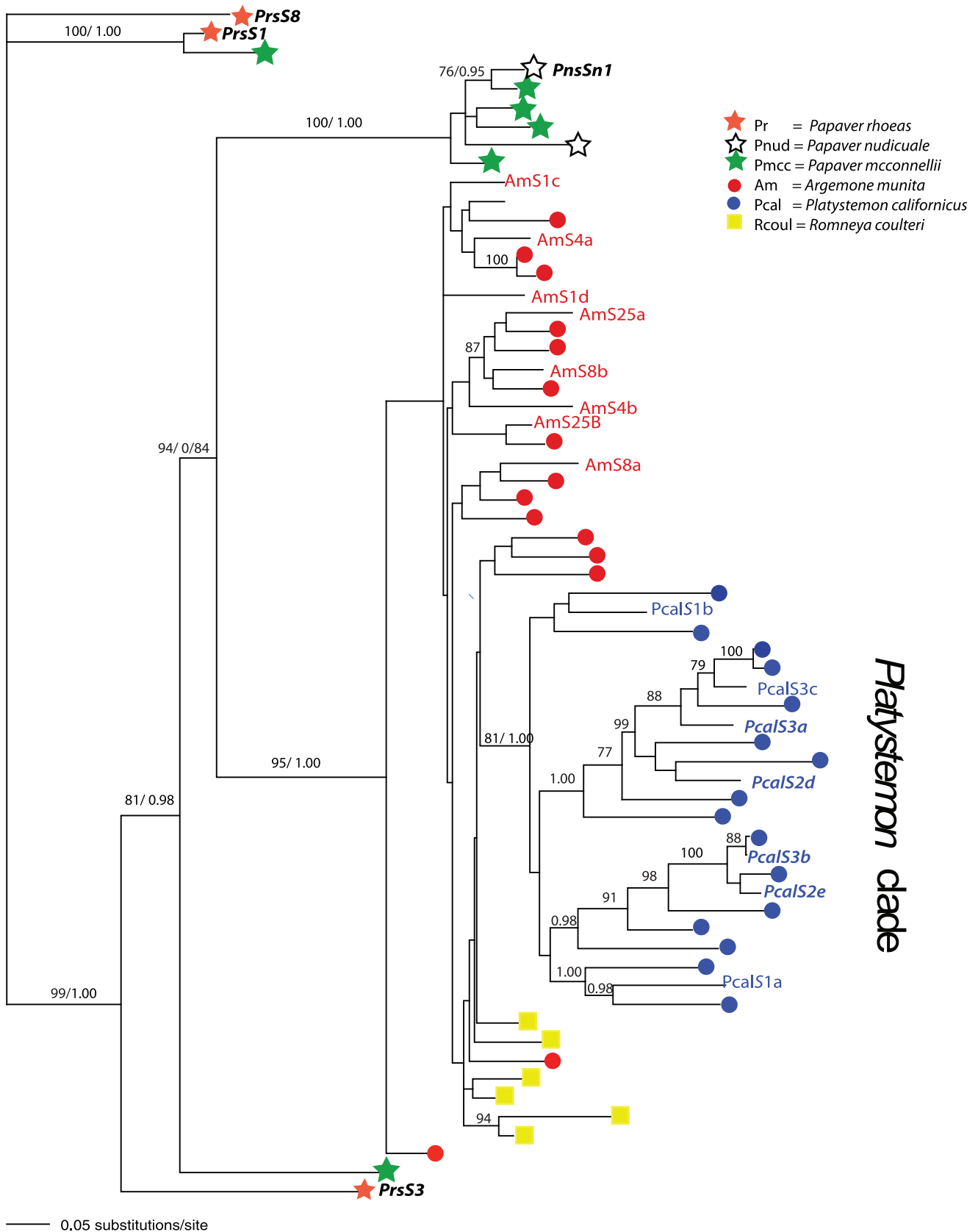doi:10.1371/journal.pone.0023635.g002

Figure 3. Maximum likelihood phylogeny of known *Papaver S*-alleles and putative *S*-alleles from other taxa. The *Papaver S*-alleles (alleles *PrsS1*, *PrsS3 PrsS8* and *PnSn1* in bold) identified by [5] and [6] form highly divergent lineages that include recently identified sequences from *P. mcconnellii* but show no close associations with any non-*Papaver* sequences. All sequences from *P. californicus* form a distinct monophyletic clade

while sequences from *A. munita* and *R. coulteri* show some poorly supported shared polymorphism. *P. californicus* linked sequences *S2d* and *S2e* from parental plant 2 are found in separate sub-clades as are co-inherited sequences *S3a* and *S3b* from parent 3 (italics). Sequences used in crosses for *A. munita* and *P. californicus* are indicated as text.
doi:10.1371/journal.pone.0023635.g003

specific polymorphism with both *P. rhoeas* and *P. nudicuale*, but nevertheless cluster together with congeneric sequences rather than with sequences derived from other genera.

Sequences from *P. californicus* form a distinct, well-supported clade and show no evidence of shared polymorphism with other taxa (Figure 3). Linked sequences *S2d* and *S2e* from the two crossing families are found in separate sub-clades within the *P. californicus* lineage as are the linked homologs *S3a* and S3b. *A. munita* and *R. coulteri* sequences show some possible shared ancestral polymorphism but these lineages are not well resolved. Overall, there is little phylogenetic structure among sequences from *A. munita* compared to those from *P. californicus*. A phylogeny including *Arabidopsis thaliana* S-protein homologs (SPHs) shows that the *A. thaliana* SPH sequences form a distinct clade relative to all Papaveraceae sequences (Figure S1). This suggests that our putative S-alleles are most closely related to functional *P. rhoeas* S-alleles and unlikely to be orthologous to any other known members of the SPH gene-family.

Sequence variability

Average pairwise estimates of synonymous, non-synonymous and total nucleotide divergence (π) show that the previously sequenced S-alleles from *Papaver* are the most variable (Table 3). Average divergence in *Papaver* is reduced slightly when sequences from *P. mcconnellii* are included due to the clustering of several sequences from this species with the previously published allele from *P. nudicaule* (see Figure 3). Sliding window analysis of 5′ sequence alignments show variation at similar positions across taxa but the amount of variation differs, with a region between 50–100 bp in *Papaver* showing the greatest diversity (Figure 4). Average pairwise diversity for the *P. californicus* sample is greater overall than for *A. munita*.

Estimates of selection

Putative S-allele sequences appear to be under positive selection. Likelihood ratio tests show that models with positive selection fit the observed data better than neutral models (Table 4). This holds for the genera with lower sequence diversity even after *Papaver* spp. sequences are removed. The empirical Bayesian method [33] identified only a few sites with statistically significant evidence for positive selection, while the majority of sites are found to be under purifying selection (Table 4). Using sequences with complete 5′ ends, datasets including and excluding *Papaver* sequences were analyzed to determine their effects on estimates of selection in *A.*

munita and *P. californicus*. Neutral null models M1 and M7 that do not allow for positive selection were rejected using likelihood ratio tests against the alternative models M2 (two rate model), M3 (three discrete rates model), and M8 (beta distribution+ω model). Posterior probability scores indicate sites 53, 107 and 119 are under selection under models M2 and M8. The discrete M3 model shows no sites under positive selection when *Papaver* is present but that 6 sites (sites 53, 77, 83, 103, 107, and 119) have high posterior scores when those sequences are excluded. Model M8 shows higher mean dN/dS when *Papaver* is excluded suggesting those highly divergent sequences may have a dampening effect on estimating sites under selection.

We also estimated selection using OmegaMap [34] because this method accounts for recombination (see Text S1 for analysis of recombination). Using sequences from *Papaver*, *Argemone* and *Platystemon*, this method finds more sites under selection (Figure 4). Codons 13, 33 and 68 were predicted to be under positive selection with OmegaMap but not PAML. Codons 107 and 119 were consistently predicted to be subject to positive selection using both methods, while codon 103 was predicted to be under positive selection in PAML model M3 and OmegaMap. Using OmegaMap, site 53 only reached high posterior probability (0.92) of positive selection after exclusion of *Papaver* sequences. It appears that the independent model of rate variation may lack power when only *A. munita* and *P. californicus* are included (22 sequences), as fewer sites are identified than when *Papaver* is included.

Discussion

For all species examined, we amplified polymorphic sequences that resemble the four previously described S-alleles from *P. rhoeas* and *P. nudicaule* [5,6]. Crossing experiments show that putative S-locus genotypes significantly predict fruit set in crosses among full-sibs of both *A. munita* and *P. californicus*, even in genotypes with co-segregating S-like sequences. In both species, crosses between parents with at least one non-matching allele nearly always set fruit, while crosses between individuals with matching putative S-locus genotypes displayed much lower rates of fruit set. In *A. munita*, fruit set following crosses among individuals with matching genotypes only occurred when allele *S25b* was present, suggesting that this may be a weak or non-functional S-allele. For *P. californicus*, although we cannot rule out leaky or partially expressed self-incompatibility [35–37], the majority of unexpected fruit set

Table 3. Average pairwise nucleotide divergence among sequences within genera.

	Synonymous (Π_s)	Non-Synonymous (Π_a)	All Sites (Π)
Papaver (S1, S3, S8, Sn1)	0.70	0.28	0.36
Papaver (all)	0.57	0.24	0.30
A. munita	0.30	0.14	0.17
P. californicus	0.49	0.20	0.26
R. coulteri	0.20	0.11	0.13

Only sequences with 5′ ends were included except *R. coulteri* where only 3′ end sequences were amplified. The four known *Papaver* S-alleles have the greatest overall divergence for both types of sites.
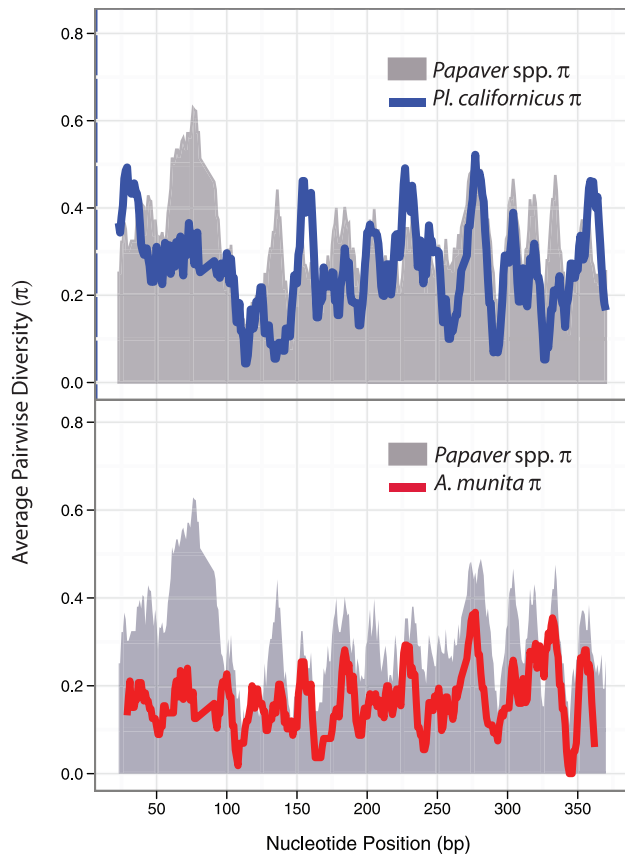doi:10.1371/journal.pone.0023635.t003

Figure 4. Sliding window analysis of average pairwise nucleotide diversity (π) of putative _S_-alleles. Data consist of 10 _Papaver_ spp (gray), 11 _P. californicus_ (blue), and 11 _A. munita_ (red) sequences possessing 5′ ends. Regions of high divergence across the three genera appear in similar positions. Divergence was estimated using DNASP [72].
doi:10.1371/journal.pone.0023635.g004

occurred in crosses among ambiguously genotyped individuals. In sum, sequences identified here and examined in crosses appear to either be functional _S_-alleles, or are linked to the functional genes, a distinction no crossing experiment can settle.

While two homologous products were found for several individuals of _A. munita_ and _P. californicus_, as expected of an obligately heterozygous GSI locus, only one product was found for several individuals of these species and for all individuals of _Papaver_. The finding of only one _S_-like sequence in some plants may imply that amplification is inconsistent among genotypes, perhaps due to PCR competition among alleles, or that other lineages of _S_-alleles exist that are too divergent to be amplified using our degenerate forward or reverse primers. Despite crossing data to show that _P. rhoeas_ possesses more than 60 _S_-alleles [4], Kurup et al. [6] were only able to sequence two alleles from _Papaver_ using various hybridization and PCR techniques to add to the two previously described [5]. However, even in systems where the range of variation at the _S_-locus is well-characterized, amplification of fewer than two alleles from individuals is common in population studies [38,39].

More than two sequences were found in some individuals of both _A. munita_ (1 of 23 plants) and _P. californicus_ (4 of 25 plants). In _A. munita_ finding more than two _S_-like sequences in at least some individuals was not entirely surprising as this species is tetraploid. However, our crossing data so far find no indication of more than four mating types among full sibs suggesting one homologue of the _S_-locus may be deleted or non-functionalized in many individuals. The "extra" copy may have been amplified in the individual from which three sequences were recovered. While polyploidy, or even duplication of just the pollen-determining part of the _S_-locus causes loss of self-incompatibility in RNase-based SI of Solanaceae and Plantaginaceae, [40–42], this finding cannot be generalized to other GSI systems [43].

In diploid _P. californicus_, finding more than two copies of _S_-like genes was unexpected. Because _P. californicus_ carries the chromosome complement considered basal in the family, it is unlikely to represent a recently diploidized polyploid. Rather, presence of multiple _S_-like sequences likely represents tandem duplication

Table 4. Likelihood ratio test statistics ($-2\Delta\text{lnL}$) compare models with selection (boldface) to preceding neutral models on 5′ end containing sequences from _Papaver_, _Argemone_ and _Platystemon_.

a)

Model	Likelihood	$-2\Delta\text{lnL}$	_p_-value (df)	Selected sites
M1	−6410.82			N/A
M2	−6406.15	9.34	_p_<0.01 (2)	107, 119
M3	−6374.01	73.62	_p_<<0.01 (3)	None
M7	−6370.63			N/A
M8	−6366.94	7.38	_p_<0.05 (2)	107*, 119

b)

Model	Likelihood	$-2\Delta\text{lnL}$	_p_-value (df)	Selected sites
M1	−4055.71			N/A
M2	−4049.04	13.34	_p_<0.01 (2)	53, 107*, 119
M3	−4034.8	41.82	_p_<<0.01 (3)	53, 77, 83, 103, 107, 119
M7	−4037.95			N/A
M8	−4030.48	14.94	_p_<<0.01 (2)	53, 107*, 119

Posterior probabilities for sites under selection are ≥0.95 except when followed by an asterisk, in which case 0.90≤p≤0.95.
5′ Alignment of 10 _Papaver_, 11 _A. munita_ and 11 _P.calfornicus_ sequences (a). 5′Alignment of 11 _A. munita_ and 11 _P.californicus_ sequences only (b).
doi:10.1371/journal.pone.0023635.t004

following divergence from *Papaver*. No linked or unlinked homologous *S*-like products have been detected in *P. rhoeas* (V.E. Franklin-Tong *personal communication*), despite considerable efforts to characterize its *S*-locus [44]. RT-PCR of RNA from *P. nudicaule* stigmas using primers designed to amplify sequences found in *A. munita* and *P. californicus* did not yield any sequences in *Papaver* closely similar to those from *A. munita*, *R. coulteri*, or *P. californicus* (data not shown). It is also unlikely that any of our sequences are orthologs of the *Arabidopsis* SPH proteins because phylogenetic estimates (Figure S1) indicate substantially closer homology to the demonstrated *Papaver* *S*-alleles.

Duplication of *S*-locus genes is apparently common in SI systems. In the sporophytic system of the Brassicaceae, homologous polymorphic stigmatic *S*-locus genes *SRK* (*S-receptor kinase*) and *SLG* (*S*-locus glycoprotein) are found in *Brassica oleracea*, *B. campestris* and *Raphanus sativas* [16,45,46] while only *SRK* is found in *Arabidopsis lyrata*, *A. halleri* and *Capsella grandiflora* [22,25,26,47–49]. *SLG* was either lost in some ancestor of both *Arabidopsis* and *Capsella* or gained in *Brassica/Raphanus*. Recently, Busch et al. [50] reported several loci in *Leavenworthia alabamica* that are similar to *SRK*, only one of which is shown to co-segregate with SI. Duplications are also abundant in the *S-RNase* system where multiple pollen-*S*-like F-Box genes are found linked to the *S*-locus in some taxa [51–53]. The presence of multiple linked loci makes it difficult to isolate the genes directly involved in SI [54–55], and recent evidence [56] strongly suggests that, at least in some cases, linked loci may act jointly to determine specificity.

The expectation for loci under balancing selection is that polymorphism will be preserved over very long time periods resulting in trans-specific polymorphism. This seems to be true of the sequences from *Papaver* which display several diverse and apparently old lineages. However, for the other taxa sampled, our phylogenetic analysis indicates a general clustering of putative *S*-alleles according to genera, relatively little phylogenetic depth to the sequences from each genus, and the lack of trans-generic lineages with the only minor exceptions being among poorly supported nodes subtending sequences from *A. munita* and *R. coulteri*.

There several potential reasons why shared ancestral polymorphism is largely absent from the sequences found here. First, the primers used in our search for S-alleles were based on the four published allele sequences from *Papaver*. Additional sequence diversity may exist within *Papaver* which could explain why studies prior to ours failed to amplify additional *Papaver* S-alleles from *P. rhoeas*. It could be that S-locus diversity outside that known from *Papaver* exists in the taxa studied here and its discovery could alter our view of the extent and distribution of ancient polymorphism. Nevertheless, the level of polymorphism in sequences amplifiable with existing primers appears considerably smaller outside of the genus *Papaver* than within it.

Second, the sequences outside of *Papaver* might not represent functional *S*-alleles, but rather paralogs of the *S*-locus not directly involved in incompatibility. The strength of balancing selection acting on such paralogs, even if linked to the *S*-locus as found here, would be much reduced relative to functional *S*-alleles, perhaps explaining their reduced levels of divergence, shorter coalescence times, and lack of shared ancestral polymorphism.

Third, lack of shared ancestral polymorphism may reflect long divergence times among the genera examined (Table S2), with corresponding turnover among the alleles maintained in each genus. Castric and Vekemans [48] highlighted a nearly monophyletic clade of *Brassica oleracea* *SRK* alleles with only two trans-specific lineages shared with *Arabidopsis lyrata* and none with *A. halleri*. This was explained by the 15–20 million year divergence of

B. oleracea from *Arabidopsis* [57] resulting in lineage sorting. Among families with RNase-based GSI, contrasting patterns are seen. In the Solanaceae, despite an estimated 40 million years of divergence among sampled genera [23], ancestral polymorphism is widely shared. In the Rosaceae, reciprocally monophyletic sets of alleles are observed from subfamilies Prunoideae and Maloideae [58] thought to have diverged some 32 MYA [59]. Minimum divergence time estimates among the genera examined here (10–40 MY; Table S2) appear at least as long as those between SSI *Arabidopsis* and *Brassica*. Following lineage sorting some variation in observed coalescence times of sequences from different species is expected, though the variation seen here between, for example, *Argemone* and *Papaver* appears extreme.

Fourth, reduced long-term population sizes and/or population bottlenecks can increase the rate of turnover affecting both levels of shared ancestral polymorphism and sequence variation found in loci under diversifying selection. For *S*-alleles, two cases of ancestral restrictions of *S*-locus diversity have been described [21,23,24]. However, the single monophyletic clade of alleles from *P. californicus*, and the nearly monophyletic lineage from *A. munita* are unlikely to be attributable to past bottlenecks. At a gametophytic *S*-locus, a self-incompatible population cannot persist with fewer than three alleles [60] so bottlenecks resulting in a single surviving lineage would be unexpected. In addition, one would have to invoke separate bottlenecks, or a bottleneck followed by turnover, to explain the reciprocally monophyletic sequences from *Platystemon* and *Argemone/Romneya*.

Finally, gene duplication and recombination/gene conversion could also cause the observed phylogenetic patterns. The expected consequence of recombination on genealogies of balanced polymorphisms is shorter times to coalescence and long terminal branches relative to depth, resulting in a topology with a star-like appearance and reduced phylogenetic structure [60–63]. Short coalescence time and lack of structure is particularly apparent in the sequences from *A. munita* suggesting that recombination might be particularly high in this taxon. While it is difficult to provide definitive tests for recombination when multiple homologous loci are present, preliminary analyses (Table S3, Figure S2) provide estimates of increased recombination in *A. munita* relative to *P. californicus*. A higher population recombination rate may explain the low phylogenetic resolution of sequences from *A. munita* relative to *P. californicus* while gene duplication and gene conversion in both of these taxa may explain both their low levels of variation and monophyletic or near-monophyletic clusters of sequences relative to *Papaver*. Using the *Brassica SRK-SLG* system as a model, Takuno *et al.* [64] demonstrated clear evidence for interlocus gene conversion. Simulations showed that when a functional SI gene is under diversifying selection, a duplicated copy can be advantageous if it contributes to variation in the functional gene through gene conversion.

Preliminary codon based estimates of selection (Table 4, Figure 5) suggest substantial constraint with several residues under positive selection. The secondary structures of the *Papaver* stigmatic *S*-proteins *PrsS1*, *PrsS3*, *PrsS8* and *PnudSn1* [14] predicted seven hydrophilic surface loops. Substitution at a single variable site within loop 6 of *PrsS1* caused loss of function. This amino acid corresponds to site 107 in our alignment (Figure 1) and attains high posterior probability of positive selection in all analyses. This functional assay counters the criticism that the types of statistical methods used here are unreliable in detecting important functional regions undergoing diversifying selection [65–67].

We have presented putative stigmatic *S*-locus sequences from several naturally occurring species of Papaveraceae. It is not yet proven that any of these sequences function directly in self-pollen
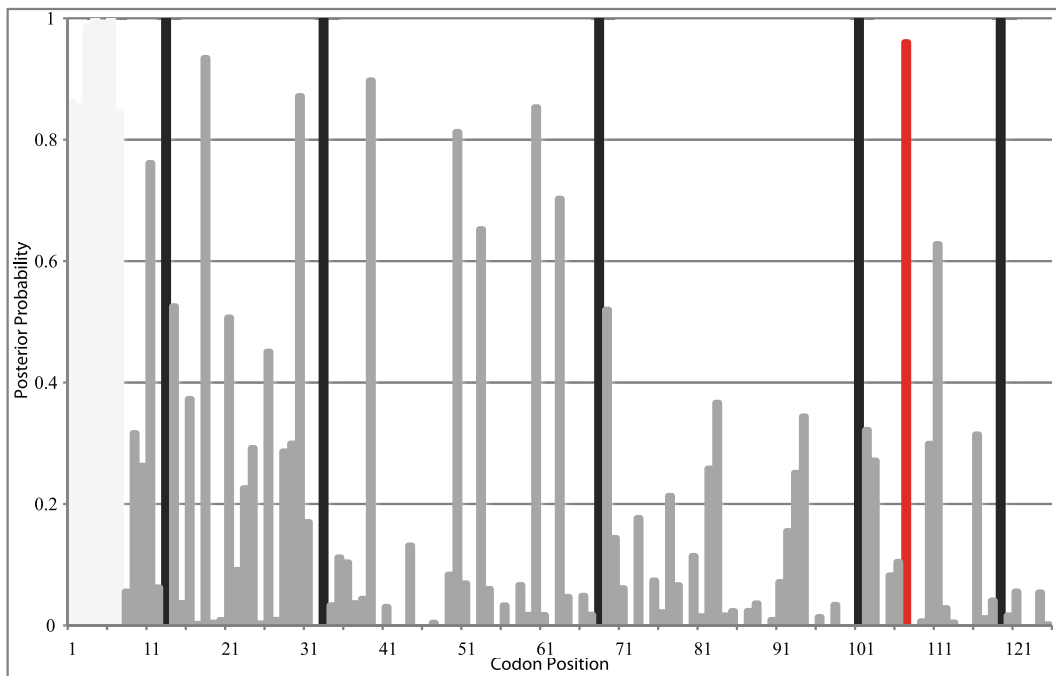
Figure 5. Posterior probabilities for codons estmatied to be under positive selection using OmegaMap [34] implemented with an independent model of rate variation for each site. The dataset used includes 10 *Papaver*, 11 *P. californicus* and 11 *A. munita* sequences. Sites 1–7 from *A. munita* and *P. californicus* are excluded because they are not present in *Papaver*. Codons 13, 33, 68, 103, 119 (black bars) and 107 (red) are predicted to be under positive selection.
doi:10.1371/journal.pone.0023635.g005

recognition and rejection. However, their homology to the known *S*-alleles of *P. rhoeas* and the correspondence between genotype and incompatibility phenotype in our test crosses suggest that many may function as S-alleles or, at the very least, are linked to the *S*-locus in these taxa. Future studies that use *in vitro* assays of the effects of expressed proteins on pollen tube inhibition [5,6] are needed to unambiguously determine function and identify potentially non-functional homologs. Preliminary sequence analyses also show evidence of positive selection, as expected for S-alleles. The phylogenetic relationships, with much lower levels of sequence diversity displayed by *A. munita* and *P. californicus* relative to the levels of polymorphism among confirmed *S*-alleles from *Papaver* are somewhat surprising, though not unprecedented among loci under balancing selection. If observed levels of polymorphism represent real differences at the *S*-locus among the genera reported here, then molecular details of the *S*-locus, and the genome in which it resides, for instance whether or not there are duplicate copies of *S*-locus genes, may play a role in shaping patterns of polymorphism at loci under balancing selection.

Materials and Methods

Plant Material

Stigmatic tissue of *P. californicus* was collected from randomly sampled individuals from three previously documented populations ([30] and G. Hannan *personal communication*): Lake Cuyamaca State Park east of San Diego (N 32°58′59.1″, W 116°33′45.4″) Hastings Natural History Reservation (N 36°23′07.1″, W 121°33′16.6″; UC Reserve Permit: 5870), and Carmel Valley Road near Carmel, CA (N 36°26′34.5″, W 121°38′53.7″). Seeds of *P. californicus* were also collected from the Hastings Natural History Reservation and germinated for greenhouse crossing experiments.

Argemone munita stigmatic tissue and seeds were collected near Valentine Eastern Sierra Reserve, Mammoth Lakes (N 37°36′51″, W 118°49′47″; UC Reserve Permit: 6370), CA and Emerson Oaks UC Reserve near Temecula, CA (N 33.466668, W -117.03944; UC Reserve Permit: 6723). *Romneya coulteri* stigmas and leaves were collected from plants growing in Del Mar, CA, Otay Mesa near San Diego, and on the UCSD campus. *P. nudicaule* (Iceland Poppy) was obtained from a commercial seed packet. *P. mcconnellii* was collected from a steep rocky scree outcrop near Highway Pass in Central Alaska (approx. N 63°28′, W 150°10′).

Bulk seed of *A. munita* collected from wild populations from San Diego County was used for greenhouse crosses. Seeds were sown individually in 1″ plug trays in moist SunGro Professional mix. Germination rates were very low for both *A. munita* and *P. californicus*, which typically required a minimum of 45–90 day cold treatment (4°C) before any seedlings emerged. *A. munita* was grown from seedlings in a glasshouse with natural light until flowering when hand pollinations were conducted. After germination, *P. californicus* was grown to flowering under fluorescent lights at 14/10 hr day/night light regimes. Full-sib families of both *A. munita* and *P. californicus* were produced from genotyped parental plants. Seeds of *P. mcconnellii* were cold stratified (4°C with 6 hr light) for one month to induce germination, and plants were grown in 1:1:1 perlite:vermiculite:coconut coir in a glasshouse with natural light.

For *A. munita* and *P. californicus*, hand pollinations among sets of full-sib offspring were used to test the relationship between putative *S*-locus genotype and incompatibility phenotype. Flowers of both species were bagged prior to anthesis to prevent pollen contamination. Replicate pollinations of three or more crosses between each reciprocal pair were attempted, though not always feasible due to limited flowers. Pollinations were made by collecting anthers using a forceps, and depositing pollen on the stigma. Crosses were made among individuals with matching sets

of sequences ('incompatible' treatment) and those that differed at a single putative allele ('semi-compatible' treatment). Self and non-matching genotype crosses were conducted in similar fashion.

The finding of more than two S-like sequences from single plants (see results) presented the possibility of a multi-locus system of self-recognition. To test for this we conducted ten 5×5 reciprocal diallel crosses (four in *A. munita* and six in *P. californicus*) among full-sib offspring. In a single-locus system, only four mating types are expected among full sibs so diallel crosses among five full siblings should always uncover at least one incompatible pair. In multi-locus systems, the number of mating types among full sibs is expected to be greater, 16 if there are two heterozygous loci which both act to confer phenotype and so diallels among five sibs might often uncover no incompatible pairs. These diallel crosses were conducted in the absence of genotypic data for the individuals used.

Molecular Techniques

Amplification of putative S-alleles was performed using reverse-transcriptase (RT-) PCR from total RNA extracted from stylar/stigmatic tissue. RNA was extracted using Trizol (Invitrogen Corp.) or the PureScript kit (Gentra Systems Inc., Minneapolis, MN). For *Papaver* spp. some sequences were amplified from genomic DNA extracted from leaf tissues with the CTAB method [68] or with PUREGENE DNA purification kit (Gentra Systems Inc., Minneapolis, MN).

The forward primers Prho1F: 5′ATG/AAY/MRR/MGR/GGN/AAY/GG and Prho4F: GTG/CGH/ATA/ATG/AAY/ARR/AGA/GG were designed from a conserved amino acid region based on the four published *Papaver* sequences and annealed to a region approximately 90 bp downstream from 5′ end. These were used in conjunction with an oligo-dT reverse primer (AUAP, Invitrogen Corp.) for amplification of 3′ ends of putative S-alleles from cDNA. Nested PCR was performed for several individuals of *P. californicus* due to non-specific amplification using the forward and reverse degenerate primers Prho1F and PopR3: GCC/ASR/TRH/ADV/ASC/WGC/C. Amplified PCR products were then sub-cloned into competent *E. coli* cells using the TOPO 2.1 TA cloning kit (Invitrogen Corp.). Restriction fragment analysis was used to identify unique products from clones.

Because there is an approximately 90 bp region upstream from the forward degenerate primers in the 4 published *Papaver* sequences, 5′ rapid amplification of cDNA ends (5′ RACE) was performed using the Roche 2nd Generation 5′ RACE kit (Roche Diagnostics). RNA was first transcribed into cDNA using allele-specific reverse primers obtained from 3′ end sequence data. This technique used the enzyme terminal transferase to ligate a poly-A region to the 5′ end of cDNA and then one or two nested PCR reactions using additional allele-specific reverse primers. To obtain full-length *Papaver* spp. sequences, genomic DNA was fragmented with a restriction digest, and an adapter was attached to the ends. PCR was conducted with an allele specific primer and a generic primer on the adapter.

To assess tissue specificity of expression of putative S-alleles, RNA was isolated from stigmas, leaves, stems and buds of unopened flowers. Total RNA was treated with Ambion Turbo DNase (Ambion Inc.) according to the supplier's protocol to remove genomic DNA. Extractions were tested for DNA contamination by performing allele specific PCR prior to reverse-transcribing to cDNA. Because it is not possible to completely remove all genomic DNA, we used allele specific forward primers and a reverse abridged universal amplification primer that anneals downstream of the poly-A tail on samples suspected to have remaining DNA. The quality of cDNA was then tested by PCR using *actin* primers.

Sequence Analysis and Phylogenetics

Sequences were aligned with BioEdit v7.05 (Ibis Therapeutics, Inc.) using the known *P. rhoeas* stigmatic S-alleles as a guide. A total of 64 sequences were used to construct the phylogeny including sequences with 5′ and 3′ ends and those with only 3′ ends. A maximum likelihood (ML) phylogenetic tree was constructed using PAUP [69] under the GTR+I+G substitution model. ModelTest 3.7 [70] was used to determine base frequencies (A = 0.314 C = 0.153 G = 0.244 T = 0.287) and the appropriate substitution model. Branch support for the ML tree was obtained by bootstrap analysis of 1000 replicates using the same base frequencies as above. Mr. BAYES [71] was also used to confirm the maximum likelihood topology and branch support. The Bayesian analysis was run using a GTR+I+G substitution model across sites for 1,000,000 generations, sampling every 100th tree for a total of 10,000 trees. A burn-in of 2500 trees was discarded and posterior probabilities were calculated on the remaining 7500 trees. No root was specified for either ML or Bayesian analysis but the *P. rhoeas* S1 and S8 alleles were found to form distant sister group to all other alleles. Since the ML and Bayesian methods produced similar topologies, only the ML tree is reported.

Sequence diversity

Average pairwise nucleotide divergence was estimated using DNASP [72]. Sliding window estimates of interspecific divergence were estimated for alignments of full-length sequences (including the 5′ region). The 5′ dataset included 10 full-length *Papaver*, 11 *P. californicus* and 11 *A. munita* sequences including 5′ ends, but with incomplete data for approximately 30 nucleotides at the 3′ end.

Estimates of positive selection

To estimate the ratio (ω) of non-synonymous (d_N) to synonymous (d_S) substitution rates at each amino acid position we used the program *codeml* in PAML 3.15 [72]. Values of $\omega < 1$ for individual codons indicate purifying selection while sites with $\omega = 1$ are considered neutral. Positive selection at the amino acid level is inferred when $\omega > 1$. Likelihood ratio tests are used to compare a series of nested neutral and selection models [32], [73] to determine the model that best fits the data. The null model M1 (neutral) constrains all sites to be either of class $\omega = 0$ or $\omega = 1$ while the alternative model M2 adds a third class in which ω is free to vary at individual sites. Model M3 assumes three discrete site classes ($\omega_0 < 1$, $\omega_1 = 1$ and $\omega_2 > 1$) with three corresponding proportions (p_0, p_1, p_2) estimated from the data. Two similar models, M7 and M8 (neutral) and the alternative M8 (selection) assume a beta distribution of rates among 11 site classes. Sites estimated to be under positive selection are determined by an empirical Bayesian approach [73] where posterior probabilities are estimated from rates within each site class. Because individual species possess several polymorphic alleles, comparisons of d_N/d_S ratios can be made within and between species and can also be estimated cumulatively [32,74–76].

We also used OmegaMap [33] to determine whether estimates of selection on individual codons differ using a coalescent method that allows for recombination. An independent model for ω at each site was used for 5′ datasets, as these were sufficiently small (32 and 22 sequences, with and without *Papaver* spp., respectively, to allow for this computationally intensive MCMC chain of 500,000 generations. An improper inverse prior distribution of ω was set for the independent model.

Supporting Information

Figure S1 Bayesian phylogeny of *Papaver* S-alleles (*PrS1*, *PrS3*, *PrS8*, and *PnSn1*) and 28 putative Papaveraceae *S*-alleles showing their relationship to 6 *Arabidopsis thaliana* S-protein homologues (SPH's). The SPH's were chosen using a maximum E-value of 10^{-4} in tBLASTx searches against the *A. thaliana* protein database. Numbers on nodes are posterior probability scores. The phylogeny is midpoint rooted but this did not effect distances between sequences. *Argemone munita* sequence S25a queried against the Flowering Plants database resulted in E-values ranging from $9e^{-16}$ to $5e^{-20}$ with closest hits to the four *Papaver* S-alleles. Similar results are obtained using any other sequence from our sample. SPH sequences were obtained from NCBI accessions 18390419, 18414536, 18416755, 22328708, 21404932 and 22328912.
(DOC)

Figure S2 Coalescent simulation of the effect of sequences from a paralogous gene on recombination detection. Some putative S-alleles in our data set may belong to one or more duplicated, paralogous S-pseudogene. These are not expected to be under balancing selection and are likely to be released from recombination suppression. Therefore, even if functional *S*-alleles experience complete recombination suppression, contamination of the dataset by recombining *S*-pseudogenes could cause statistically detectable recombination. The goal of this simulation was to evaluate the effect of *S*-pseudogene haplotypes on the detection of recombination in a combined data set. We conducted coalescent simulations with Hudson's [77] ms program, modified to allow changes in recombination rate over time. The allelic genealogy of S-alleles can be approximated with a neutral coalescent with a scaling factor, f [78]. We assumed that the population parameter, θ_s, of S-alleles to be $4f N_e \mu$, where N_e is the neutral effective population size and μ is a neutral mutation rate. For the paralogous pseudo-gene, which has been released from balancing selection, the population parameter was set to that of a neutral gene ($\theta_p = 4 N_e \mu$). At t_d generations ago (in units of $4 N_e f$), there was a duplication of an S-allele and a paralogous pseudo-gene was created. We assumed that functional S-alleles do not recombine (the gray area), but the pseudo gene is released from recombination suppression and recombination within the paralogous alleles can occur immediately after the gene duplication (black region) with the recombination rate between the neighboring nucleotide sites of r per generation. Since we assumed that the pseudo gene was created by a single duplication event, we forced the S-pseudogene to have a severe bottleneck of θ_b for t_b generations (in units of $4 N_e f$). In the coalescent simulation, we assumed that we have sampled 24 sequences, of which some proportion, c (25%, 50%, or 75%), is actually contamination from the S-pseudogene. For each genealogical tree created by the coalescent simulation, DNA sequence evolution was simulated by Seq-Gen version 1.3.2 [79] under the Jukes-Cantor model [80], the non-parametric permutation tests of recombination were conducted with permute (r^2 statistics) in OmegaMap v0.5 [33] and we recorded the proportion of simulations with the p-value≤0.05. We assumed a sequence length of 327 bp, which is close to that of our data, and the following parameters were used: $f = 20$; θ_s per site = 0.01, 0.1, 0.2, 0.3, 1, or 2; $t_d = 0.01$, 0.1, 0.1, 1, 2, or 3; $r/\mu = 2$, 20, 60, 100, 200, or 2000; $t_b = 0.005$; θ_b per site = θ_s/2000000. All factorial combinations of the parameter values were used, and 1000 simulations were run for each set. In our simulations, the probability of detecting recombination with the nonparametric permutation did not exceed 0.075. This is because compared to the large diversity of S-alleles, the newly duplicated pseudo-gene has a low diversity and recombination within the pseudogene does not contribute much to the detection of recombination. These results suggest that it is unlikely that the recombination detected in our dataset is due to inclusion of S-pseudogenes.
(DOC)

Table S1 Segregation of parental alleles among full sibs used for testing the relationship between putative SI genotype and phenotype. Tests compare expectations for two alleles from each parent segregating separately to create four offspring genotypes. For crosses involving *Pl. californicus*, two sequences from one parent were always found together in offspring or else neither was found. These are considered linked polymorphisms (21a,b) and (26a,b). From plant 26, an alternative allele (26c) was recovered in offspring where 26a,b was absent (see main text).
(DOC)

Table S2 Estimated divergence times among taxa. Divergence times were estimated as the number of silent substitutions per site divided by the silent substitution rate (substitutions per site per year). Three genes were used for these estimates: the ribosomal internal transcribed spacer (ITS) and chloroplast genes (*rbc*L and *atp*B). Sequences were obtained from NCBI (AF057672, AF098922, AF305331, DQ912885, U86621, U86632, U86630, FJ626614, U86393, U86394, U86395, U86399), and the number of silent substitutions was estimated from DNASP [69]. Several estimates of angiosperm silent substitution rates were obtained for each gene [81–83], and the most extreme values were used for the min and max estimates of Papaveraceae divergence times.
(DOC)

Table S3 Non-parametric permutation analyses of the correlation between linkage disequilibrium and distance (top) and coalescent likelihood estimates of the population mutation ($\theta = 4N_e\mu$) and recombination ($\rho = 4N_e r$) rates with 95% highest posterior densities (bottom).
(DOC)

Text S1 Methods describing estimation of recombination rates in *A. munita* and *P. californicus*.
(DOC)

Acknowledgments

Author Contributions

Conceived and designed the experiments: TP JK. Performed the experiments: TP TN. Analyzed the data: TP DW NT JK. Contributed reagents/materials/analysis tools: DW NT JK. Wrote the paper: TP JK.

References

1. de Nettancourt D (1977) Incompatibility in angiosperms. Berlin: Springer. 230 p.

2. Igic B, Lande R, Kohn JR (2008) Loss of self-incompatibility and its evolutionary consequences. Int J Plant Sci 169: 93–104.

3. Lawrence MJ, O'Donnell S (1981) The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. III. The number and frequency of S-alleles in two further natural populations. Heredity 47: 53–61.

4. Lane MD, Lawrence MJ (1993) The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VII. The number of S-alleles in the species. Heredity 71: 596–602.

5. Foote HG, Ride JP, Franklin-Tong VE, Walker EA, Lawrence MJ, et al. (1994) Cloning and expression of a distinctive class of self-incompatibility (S-) gene from *Papaver rhoeas*. Proc Natl Acad Sci U S A 91: 2265–2269.

6. Kurup S, Ride JP, Jordan N, Fletcher G, Franklin-Tong VE, et al. (1998) Identification and cloning of related self-incompatibility S-genes in *Papaver rhoeas* and *Papaver nudicaule*. Sex Plant Reprod 11: 192–198.

7. Wheeler MJ, Franklin-Tong VE, Franklin FCH (2001) The molecular and genetic basis of pollen-pistil interactions. New Phytol 128: 565–580.

8. Franklin-Tong VE, Franklin FCH (2003) Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. Trends in Plant Sciences 8: 598–605.

9. Thomas S, Osman K, deGraaf BHJ, Shevchenko G, Wheeler M, et al. (2003) Investigating mechanisms involved in the self-incompatibility response in *Papaver rhoeas*. Phil Trans Roy Soc Lond B 358: 1033–1036.

10. Thomas SG, Franklin-Tong VE (2004) Self-incompatibility triggers programmed cell death in *Papaver* pollen. Nature 429: 305–308.

11. de Graaf BHJ, Rudd JJ, Wheeler MJ, Perry RM, Bell EM, et al. (2006) Self-incompatibility in *Papaver* targets soluble inorganic pyrophosphatases in pollen. Nature 444: 490–493, 230.

12. Wheeler MJ, de Graaf BHJ, Hadjiosif N, Perry RM, Poulter NS, et al. (2009) Identification of the pollen self-incompatibility determinant in *Papaver rhoeas*. Nature 459: 992–995.

13. Walker EA, Ride JP, Kurup S, Franklin-Tong VE, Lawrence MJ, et al. (1996) Molecular analysis of two functional homologues of the S3 allele of the *Papaver rhoeas* self-incompatibility gene isolated from different populations. Plant Mol Biol 30: 983–994.

14. Kakeda K, Jordan ND, Conner A, Ride JP, Franklin-Tong VE, et al. (1998) Identification of residues in a hydrophilic loop of the *Papaver rhoeas* S protein that play a crucial role in recognition of incompatible pollen. Plant Cell 10: 1723–1732.

15. Ioerger TR, Gohlke JR, Xu B, Kao T-h (1991) Primary structural features of the self-incompatibility protein in Solanaceae. Sex Plant Reprod 4: 81–87.

16. Kusaba M, Nishio T, Satta Y, Hinata K, Ockendon D (1997) Striking sequence similarity in inter- and intra-specific comparisons of class I *SLG* alleles from *Brassica oleracea* and *Brassica campestris*: Implications for the evolution and recognition mechanism. Proc Natl Acad Sci U S A 94: 7673–7678.

17. Hughes A, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167–170.

18. Ride JP, Davies EM, Franklin FCH, Marshall DF (1999) Analysis of *Arabidopsis* genome sequence reveals a large new gene family in plants. Plant Mol Biol 39: 927–932.

19. Testa G, Caccia R, Tilesi F, Soressi GP, Mazzucato A (2002) Sequencing and characterization of tomato genes putatively involved in fruit set and early development. Sex Plant Reprod 14: 269–277.

20. Ioerger TR, Clark AG, Kao T-h (1990) Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc Natl Acad Sci U S A 87: 9732–9735.

21. Richman AD, Uyenoyama MK, Kohn JR (1996) Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. Science 273: 1212–1216.

22. Bechsgaard JS, Castric V, CharleswothD, Vekemans X, Schierup MH (2006) The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr. Mol Biol Evol 23: 1741–1750.

23. Paape T, Igic B, Smith S, Olmstead R, Bohs L, et al. (2008) A 15-million-year-old genetic bottleneck at the S-locus of the Solanaceae. Mol Biol Evol 25: 655–663.

24. Miller JS, Levin RA, Feliciano NM (2008) A tale of two continents: Baker's rule and the maintenance of self-incompatibility in *Lycium* (Solanaceae). Evolution 62: 1052–1065.

25. Guo Y, Bechsgaard JS, Slotte T, Neuffer B, Lascoux M, et al. (2009) Recent speciation of *Capsella rubella* from *Capsella grandiflora*, associated with loss of self-incompatibility and an extreme bottleneck. Proc Natl Acad Sci U S A 106: 5246–5251.

26. Foxe JP, Slotte T, Stahl EA, Neuffer B, Hurka H, et al. (2009) Recent speciation associated with the evolution of selfing in *Capsella*. Proc Natl Acad Sci U S A 106: 5241–5245.

27. Beatty AV (1936) Genetic studies on the California poppy. J Heredity 27: 330–338.

28. Cook S (1962) Genetic system, variation and adaptation in *Eschscholzia californica*. Evolution 16: 278–299.

29. Wright GM (1979) Self-incompatibility in *Eschscholzia californica*. Heredity 43: 429–431.

30. Hannan G (1981) Flower color polymorphism and pollination biology of *Platystemon californicus*. Amer J Bot 68: 233–243.

31. Bilinski P (2010) Sites of self-pollen tube inhibition in Papaveraceae (sensu lato). MS Thesis, University of California San Diego.

32. Solstad H (2008) Taxonomy and evolution of the diploid and polyploid *Papaver* sect. *Meconella* (Papaveraceae). PhD Dissertation, University of Oslo.

33. Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148: 929–936.

34. Wilson DJ, McVean G (2006) Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172: 1411–1425.

35. Good-Avila SV, Stephenson AG (2002) The inheritance of modifiers conferring self-fertility in the partially self-incompatible perennial, *Campanula rapunculoides*, L. (Campanulaceae). Evolution 56: 263–272.

36. Stone JL (2004) Sheltered load associated with S-alleles in *Solanum carolinense*. Heredity 92: 335–342.

37. Mena-Ali JI, Stephenson AG (2007) Segregation analyses of partial self-incompatibility in self and cross progeny of *Solanum carolinense* reveal a leaky S-allele. Genetics 177: 501–510.

38. Vieira CP, Charlesworth D (2002) Molecular variation at the self-incompatibility locus in natural populations of the genera *Antirrhinum* and *Misopates*. Heredity 88: 172–181.

39. Mable BK, Schierup MH, Charlesworth D (2003) Estimating the number, frequency, and dominance of S-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. Heredity 90: 422–431.

40. Golz JF, Su V, Clarke AE, Newbigin E (1999) A molecular description of mutations affecting the pollen component of the *Nicotiana alata* S locus. Genetics 152: 1123–1135.

41. Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, et al. (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. Nature 429: 302–305.

42. Qiao H, Wang F, Zhao L, Zhou J, Lai Z, et al. (2004) The F-box protein AhSLF-S2 controls the pollen function of S-RNase-based self-incompatibility. Plant Cell 16: 2307–2322.

43. Mable BK (2004) Polyploidy and self-compatibility: is there an association? New Phytol 163: 803–811.

44. Bosch M, Franklin-Tong VE (2008) Self-incompatibility in *Papaver*: signaling to trigger PCD in incompatible pollen. J Exp Bot 59: 481–490.

45. Sakamoto K, Kusaba M, Nishio T (1998) Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gene (SLR1) in *Raphanus sativus* L. and self-incompatible ornamental plants in the Brassicaceae. Mol Gen Genet 258: 397–403.

46. Miege C, Ruffio-Châble V, Schierup MH, Cabrillac D (2001) Intra-haplotype polymorphism at the *Brassica* S locus. Genetics 159: 811–822.

47. Schierup MH, Mable BK, Awadalla P, Charlesworth D (2001) Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. Genetics 158: 387–399.

48. Castric V, Vekemens X (2004) Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. Mol Ecol 13: 2873–2889.

49. Paetsch M, Mayland-Quellhorst S, Nueffer B (2006) Evolution of the self-incompatibility system in the Brassicaceae: identification of S-locus receptor kinase (SRK) in self-incompatible *Capsella grandiflora*. Heredity 97: 283–290.

50. Busch JW, Sharma J, Schoen DJ (2008) Molecular characterization of Lal2, an SRK-like gene linked to the S-Locus in the wild mustard *Leavenworthia alabamica*. Genetics 178: 2055–2067.

51. Sassa H, Kakui H, Miyamoto M, Suzuki Y, Hanada T, et al. (2007) S locus F-box brothers: Multiple and pollen-specific F-box genes with S haplotype-specific polymorphisms in apple and Japanese pear. Genetics 175: 1869–1881.

52. Zhou JL, Wang F, Ma WS, Zhang YS, Han B, et al. (2003) Structural and transcriptional analysis of S-locus F-box genes in *Antirrhinum*. Sex Plant Reprod 16: 165–177.

53. Newbigin E, Paape T, Kohn JR (2008) RNase based self-incompatibility: Puzzled by pollen S. Plant Cell 20: 2286–2292.

54. Hagenblad J, Bechsgaard J, Charlesworth D (2006) Linkage disequilibrium between incompatibility locus region genes in the plant *Arabidopsis lyrata*. Genetics 173: 1057–1073.

55. Wheeler D, Newbigin E (2007) Expression of 10 S-Class SLF-like genes in *Nicotiana alata* pollen and its implications for understanding the pollen factor of the S-locus. Genetics 177: 1–10.

56. Kubo K, Entani T, Takara A, Wang N, Fields AM, et al. (2010) Collaborative non-self recognition system in S-RNase-Based self-incompatibility. Science 330: 796–799.

57. Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, et al. (2001) Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana*. Plant Cell 13: 627–643.

58. Kohn JR (2008) What genealogies of S-alleles tell us. In: Franklin-Tong VE, ed. Self-incompatibility in flowering plants: Evolution, diversity and mechanisms. Berlin: Springer-Verlag. pp 103–121.

59. Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: Calibrating the family tree. Proc Roy Soc Lond B 268: 2211–2220.

60. Wright S (1939) The distribution of self-sterility alleles in populations. Genetics 24: 538–552.

61. Wiuf C, Hein J (2000) The coalescent with gene-conversion. Genetics 155: 451–462.

62. Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. Genetics 156: 879–891.

63. Schierup MH, Mikkelsen AM, Hein J (2001) Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. Genetics 159: 1833–184.

64. Takuno S, Nishio T, Satta Y, Innan H (2008) Preservation of a pseudogene by gene conversion and diversifying selection. Genetics 180: 517–531.

65. Hughes A (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. Heredity 99: 364–373.

66. Hughes A (2008) The origin of adaptive phenotypes. Proc Natl Acad Sci U S A 105: 13193–13194.

67. Yokoyama S, Tada T, Zhang H, Britt L (2008) Elucidation of phenotypic adaptations: Molecular analysis of dim-light vision proteins in vertebrates. Proc Natl Acad Sci U S A 105: 13480–13485.

68. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissues. Phytochem Bull 19: 11–15.

69. Swofford DL (2002) PAUP. Phylogenetic analysis using parsimony (and other methods) Version 4. Sunderland: Sinauer.

70. Posada D, Crandall KA (2002) The effect of recombination on the accuracy of phylogeny estimation. J Mol Evol 54: 396–402.

71. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17: 754–755.

72. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19: 2496–2497.

73. Yang Z (2000) Phylogenetic analysis by maximum likelihood (PAML). Comput Appl Biosci 13: 555–556.

74. Yang Z, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22: 1107–18.

75. Ikeda K, Igic B, Ushijima K, Yamane H, Hauck NR, et al. (2004) Primary structural features of the S haplotype-specific F-box protein, SFB in Prunus. Sex Plant Reprod 16: 235–243.

76. Turner LM, Hoekstra HE (2006) Adaptive evolution of fertilization proteins within a genus: variation in ZP2 and ZP3 in deer mice (Peromyscus). Mol Biol Evol 23: 1656–1669.

77. Hudson R (2002) Generating samples under a Wright-Fisher neutral model. Bioinformatics 18: 337–338.

78. Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. Proc Natl Acad Sci U S A 87: 2419–2423.

79. Rambaut A, Grassly NC (1997) Seq-Gen: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comp Appl Biosci 13: 235–238.

80. Jukes TH, Cantor CR (1969) Evolution of protein molecules. In Munro HN, ed. Mammalian protein metabolism. New York: Academic Press New York. pp 21–132.

81. Wolfe KH, Sharp PM, Li W-H (1989) Rates of synonymous substitution in plant nuclear genes. J Mol Evol 29: 208–211.

82. Doebley J, Durbin M, Golenberg EM, Clegg MT, Ma DP (1990) Evolutionary analysis of the large subunit of carboxylase (rbcL) nucleotide sequence among the grasses (Gramineae). Evolution 44: 1097–1108.

83. Kay KM, Whittall JB, Hodges SA (2006) A survey of nuclear ribosomal internal transcribed spacer substitution rates across angiosperms: an approximate molecular clock with life history effects. BMC Evol Biol 6: 36.