Appendix includes the following sections:

ESM4

***ITS2* analysis**

ITS2 is multicopy nuclear sequence. Therefore, it is expected to display two kind of polymorphism: allelic polymorphism and polymorphism between copies. Heterogeneous nucleotide positions were identified through dual/multiple peaks present in electropherograms and coded as N. In evolution of *ITS2* sequences, the mono, bi- and mullti-nucleotide insertions/deletions are frequent and contain phylogenetically important information. To account for this, each indel event was coded as a binary character (1/0, presence/absence of the gap independently of its length). Heterogeneous indel positions were identified by a sharp transition in the electropherogram from clean to garbled sequence, where the transition corresponded to the same position of a homozygous indel in other individuals; they were coded as N.

For indel event codification, the following gaps were considered (based on alignment ESM3):

1) one nucleotide insertion G (orN) in morgani (position 74)
2) one nucleotide insertion C in schuriani (position 108)
3) one nucleotide insertion C in schuriani (position 124)
4) one nucleotide insertion T in schuriani (position 127)
5) one nucleotide insertion A in peilei (position 128)
0) we did not consider gap 168-170 since its interpretation is difficult
6) tree nucleotide insertion AAA in dama (positions 233-235)
7) multi nucleotide insertion (nT) in dama/guezelmavi/theresiae/karindus/peilei (positions 325-340)
8) multi nucleotide deletion in morgani (positions 316-363)
9) one nucletide insertion A in schuriani and theresiae (position 343)
10) one nucleotide insertion C in karindus/peilei (position 344)0
11) three nucleotide insertion AAAAA in karindus (positions 349-351)
12) five nucleotide insertion AAACG in theresiae (positions 354-358)
13) three nucleotide insertion AAT in carmon/schuriani (positions 359-361).

This resulted in the following coding:

AY556558_birunii_MW00072 0000000000000

AY556622_carmon_MW98009 0000000000001

AY556640_dama_MW98205 0000011000000

AY_556651_guezelmavi_MW98294 0000001000000

AY556646_schuriani_MW98261 0111000010001

AY556645_theresiae_MW98240 0000001010010

F150_theresiae 0000001000010

AY556554_hamadanensis_MW00032 0000000000000

AY556745_morgani_Saq 0000000100000

Z788_morgani_Saq 1000000100000

Z841_morgani_Saq 1000000100000

Z671_morgani_N_Che 1000000100000

Z698_morgani_N_Che 1000000100000

Z699_morgani_N_Che 1000000100000

Z600_morgani_W_Sen 1000000100000

Z613_morgani_W_Sen 1000000100000

W154_morgani_W_Sen 1000000100000

W155_morgani_W_Sen 1000000100000

Z524_morgani_E_Sen 1000000100000

Z528_morgani_E_Sen 1000000100000

Z561_morgani_E_Sen 1000000100000

E473_morgani_Tak 1000000100000

AY556573_morgani_femininoides_MW00226 1000000100000

AY556749_morgani_femininoides_WE02671 1000000100000

V127_morgani_consensus 1000000100000

Z727_karindus_Saq 0000001000100

Z750_karindus_Saq 0000001001100

Z396_karindus_Van 0000001001100

Z397_karindus_Van 0000001001100

Z753_karindus_Saq 0000001001100

Z800_karindus_Saq 0000001001100

E398_karindus_Saq 0000001001100

Z381_karindus_Van 0000001001100

Z398_karindus_Van 0000001001100

Z704_karindus_consensus 000000100N100

V145_karindus_consensus 0000001001100

z763_peilei_Saq 0000001001000
z798_peilei_Saq 0000001001000
z831_peilei_Saq 0000001001000
z846_peilei_Saq 0000001001000
z662_peilei_N_Che 0000001001000
z675_peilei_N_Che 0000001001000
z700_peilei_N_Che 0000001001000
z702_peilei_N_Che 0000001001000
z568_peilei_E_Sen 0000001001000
z587_peilei_E_Sen 0000001001000
z588_peilei_E_Sen 0000001001000
W136_peilei_consensus 0000001001000
W202_peieli_consensus 0000101001000

We cloned *ITS2* for 5 specimens for which we were not able to obtain clear phylogenetic information by using standard sequencing. 44 clones were obtained for these 5 specimens (see ESM2), and strict consensus *ITS2* sequences for these specimens (ESM2 and ESM3) were then used for phylogenetic inference.

Bayesian trees were inferred using partitioned models: GTR for nucleotide substitutions and Standard model for indels as implemented in MrBayes 3.2 [37].

Cloning methods

1.1. Amplification of *ITS2*.

ITS2 region was amplified using the primer pair: ITS-3 and ITS-4 [63]. When ITS-3 and ITS-4 primers failed to amplify a sufficient product, self-designed lepidopteran primers were used: ILYC2F 5`- GAGAAACATCCAGGACCACT - 3` and ILYC2RB 5` - CTGATCTGAGGCCAACG - 3`. Amplified fragments were purified using GeneJET Gel Extraction Kit (Fermentas, Lithania). Purification was carried out according to the manufacturer's protocol. The success of PCR amplification and purification was evaluated by electrophoresis of the products in 1% agarose gel. Approximately 25 ng of PCR product were used for subsequent cloning.

1.2. Ligation of PCR product into pJet1.2 plasmid

Purified PCR products were checked on an agarose gel to avoid contamination with primer-dimers and cloned into blunt-end cloning vector pJET1.2 (Fermentas, Lithuania) according to the

manufacturer's protocol for 10 minutes at room temperature. The pJet1.2 plasmid selects successful ligations through the disruption of an otherwise lethal gene, eco47IR, which enables positive selection of the recombinants. Before ligation, a 3'-A overhang were removed from the PCR products by treating the PCR product with a proofreading DNA polymerase.

1.3 Transformation

5 µl of the ligation mixture reaction were added to 50 µl of chemo-competent *E. coli* DH101B cells an incubated for 10 min. on ice. After incubation transformation mixture were pipetted onto pre-warmed LB Anp IPTG agar plate and spread by using inoculation loop. Agar plates with competent *E.coli* were incubated overnight at 37°C.

1.4 Analysis of cloning results and plasmid DNA preparation.

For each cloning, more than 500 clones were obtained. To check if the cloning procedures were successful, PCR with *ITS2*-speciffic primers were conducted for 20 colonies per cloning reaction. GeneJET Plasmid Miniprep Kit (Fermentas, Lithuania) were used for preparation of plasmid DNA from recombinant *E. coli* culture. A single colony from a freshly streaked selective plate were picked to inoculate 1-5 mL of LB medium supplemented with ampicillin and incubated for 12-16 hours at 37°C while shaking at 200-250 rpm. The bacterial culture was harvested by centrifugation at 8000 rpm (6800 × g) in a microcentrifuge for 2 min at room temperature. The supernatant was decanted and all remaining medium was removed. The pelleted cells were resuspended and subjected to SDS/alkaline lysis to liberate the plasmid DNA. The resulting lysate was neutralized to create appropriate conditions for binding of plasmid DNA on the silica membrane in the spin column. Cell debris and SDS precipitate were pelleted by centrifugation, and the plasmid DNA were washed to remove contaminants and eluted.

1.6. Plasmid DNA Sequencing.

Plasmid DNA was sequenced using 3500xL analyzer (Applied Biosystems). Not less than 300 ng of plasmid DNA template was used for sequencing procedure. Cloned fragments were analyzed edited and aligned in Bioedit Software [32].

63. White, T.J., Bruns, T., Lee, S. & Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols, a Guide to Methods and Applications* (eds M.A. Innis, D.H. Gelfand, J.J. Sninsky & T.J. White), pp. 315-322. New York: Academic Press.

ESM5

AFLP analysis

The AFLP fragments were analyzed on 3500xL analyzer (Applied Biosystems). The raw data were visualized, aligned with the internal size standard Gene Scan ™ 600 LIZ ®, manually controlled and analyzed in GeneMapper 2.4.2 (Life Technologies). Only fragments between 70 and 500 base pairs were considered. Thresholds of 100 rfu (relative fluorescence unit) were set to remove instrument noise.

The mismatch error rates were scored using AFLPSCORE v.1.3b software [39]. To estimate mismatch error rate, we replicated the AFLP procedure, for 10 individuals of each species for two primer combinations: *Eco*RI-ACT/*Mse*I-CAT, *Eco*RI-ACT/*Mse*I-CTC.

Table.S5.1. The AFLP mismatch error rates scored by AFLPSCORE v.1.3b software [39]

Species	Mismatch error rate (%)	
	*Eco*RI-ACT/*Mse*I-CAT	*Eco*RI-ACT/*Mse*I-CTC
Polyommatus morgani	0.9	1.34
Polyommatus peilei	1.35	1.98
Polyommatus karindus	1.11	1.58

These error rates are comparable to those reported for AFLP data from other lycaenid butterflies [12].

Table S5.2. Results of the Bayesian assignment analysis of AFLP markers. Assignment probabilities under assumption of two (k=2) and three clusters (k=3) are shown.

Specimen	Assignment Probability k =2		Assignment Probability k =3		
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 3
karindus 1	0.023	0.977	0.016	0.941	0.043
karindus 2	0.036	0.964	0.032	0.933	0.035
karindus 3	0.013	0.987	0.000	0.971	0.029
karindus 4	0.009	0.991	0.000	0.967	0.033
karindus 5	0.021	0.979	0.000	0.966	0.034
karindus 6	0.015	0.985	0.000	0.961	0.039
karindus 7	0.008	0.992	0.000	0.959	0.041
karindus 8	0.022	0.978	0.000	0.944	0.056

karindus 9	0.031	0.969	0.020	0.932	0.048
karindus 10	0.007	0.993	0.000	0.971	0.029
peilei 1	0.454	0.546	0.025	0.038	0.937
peilei 2	0.309	0.691	0.000	0.021	0.979
peilei 3	0.582	0.418	0.034	0.024	0.942
peilei 4	0.355	0.645	0.000	0.044	0.956
peilei 5	0.418	0.582	0.000	0.020	0.980
peilei 6	0.164	0.836	0.000	0.168	0.832
peilei 7	0.421	0.579	0.000	0.123	0.877
peilei 8	0.461	0.539	0.034	0.051	0.915
peilei 9	0.405	0.595	0.000	0.082	0.918
peilei 10	0.655	0.345	0.048	0.044	0.908
morgani 1	0.973	0.027	0.973	0.000	0.027
morgani 2	0.963	0.037	0.950	0.000	0.050
morgani 3	0.900	0.100	0.923	0.000	0.077
morgani 4	0.982	0.018	0.976	0.000	0.024
morgani 5	0.891	0.109	0.876	0.044	0.080
morgani 6	0.899	0.101	0.954	0.000	0.046
morgani 7	0.955	0.045	0.958	0.000	0.042
morgani 8	0.977	0.023	0.978	0.000	0.022
morgani 9	0.918	0.082	0.925	0.038	0.037
morgani 10	0.951	0.049	0.957	0.000	0.043

ESM6.

χ^2-test for random vs. non-random distribution of the bearers of H1 haplogroup in the zone of sympatry between M, K and P

The incomplete lineage sorting scenario predicts that differentiated and non-differentiated haplotypes of K and M should be stochastically (i.e. randomly) distributed within different populations.

To check this prediction we used χ2-test for random vs. non-random distribution of the bearers of H1 haplogroup in the zone of sympatry between M, K and P.

In our case, of the 156 specimens of M, K and P studied, 90 specimens belong to the haplogroup H1 (figure 3*d*). Thus, the percentage of the H1 bearers is 90/156=57.69%.

Of these 156 specimens, 90 specimens were collected in the zone of sympatry. If H1 haplogroup was inherited from a common ancestor of M+K+P as predicted by incomplete lineage sorting scenario, it should be randomly distributed within different populations (H_0 hypothesis). Therefore, we would expect that in the sympatry zone 90 x 0.5769 = 52 specimens will be bearers of the H1 haplogroup.

In fact, all 90 specimens from the sympatry zone belong to the H1 haplogroup.

Thus, the **observed** number of the H1 bearers in the zone of sympatry is **90**.

The **expected** number of the H1 bearers in the zone of sympatry under H_0 hypothesis is **52**.

We calculated the χ^2-value:

χ^2 = (observed – expected)^2/expected= (90-52)^2/52= 27.77

The degree of freedom is 1 (df=no.categories-1=2-1=1).

The calculated χ^2 of 27.77 is larger than the tabular value of 10.83 (0.001 level of significance); therefore, we reject the H_0 hypothesis and conclude that the bearers of the H1 haplogroup are more frequent in the zone of sympatry than would be expected by chance.

Conclusion. The discovered distribution is extremely non-random with all the non-differentiated or weakly differentiated haplotypes of K and M concentrated in the zone of their sympatry. The deviation from a random distribution is highly significant ($P<0.001$ based on χ^2-test).

ESM7

Statistical test for hybridization vs. incomplete lineage sorting

The statistical test for hybridization vs. incomplete lineage sorting [44, 47] revealed low probability (P=0.039) that incomplete lineage sorting could account for the absence of divergence between h1 haplotypes of K and M (supplementary table S7).

Supplementary table S7. Results of the statistical test for hybridization vs. incomplete lineage sorting [44, 47]. Probabilities of observing the minimum *COI* distance in the species pairs according to a scenario without hybridization are shown.

Comparison	minDist	Probability
K - B	0.0355	0.469
M - B	0.0355	0.469
M - K	**0**	**0.039**
P - B	0.0448	0.711
P - K	0	0.058
P - M	0	0.363
C - B	0.0401	0.596
C - K	0.0262	0.383
C - M	0.0216	0.230
C - P	0.0278	0.388

B is *A. biruni*, C is *A. carmon*, K is *A. karindus*, M is *A. morgani*, and P is *A. peilei*.

ESM8

Evolutionary scenarios of HHS in P

Most likely, the hybridization between K and M and formation of P occurred in a locality near Saqqez. This supposition is in accordance with the facts that all the three species are found sympatrically in this locality, and only here they share the same mitochondrial haplotypes (figure 3c-d).

There are two possible scenarios: (1) this locality was for a long time inhabited by K and M, and the speciation in P was entirely sympatric; (2) speciation in P started in zone of secondary contact between K and M. In fact, the difference between them is not principal: in both scenarios, sympatry between K and M is necessary, and the difference is only in prehistory of K and M (were they sympatric for a long time in the place of hybridization, or their areas overlapped relatively recently).

ESM9

Parsimonious analysis of AFLP markers

244 AFLP markers were specific for one of the species studied (61 for K, 46 for M, and 5 for P) or for a combination of two species (11 for K+M, 62 for K+P, and 59 for M+P). These markers were considered as possible apomorphies (synapomorphies) and were used in the consequent analysis.

We mapped these markers on the trees describing different theoretically possible scenarios of speciation in this group of species (figure S9). The data were analysed using classic Hennigian approach based on distinguishing between synapomorphies and homoplasies [64, 65].

The scenarios of divergent speciation (figure S9, *a-c*) can not be explained without assumption about multiple homoplasies involved in evolution.

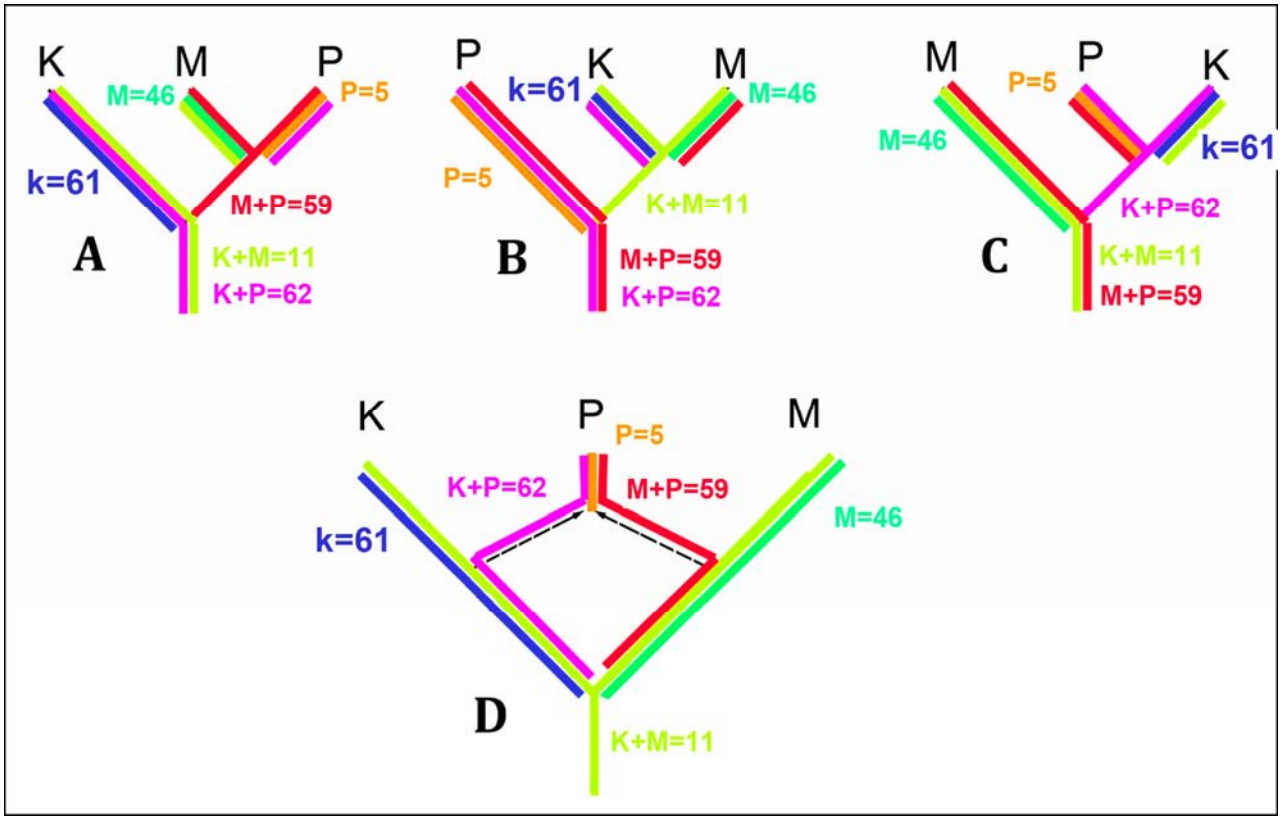


Fig. S9. AFLP markers are mapped on the trees describing different theoretically possible scenarios of speciation.
(A-C) are scenarios of divergent speciation. D is scenario of hybrid speciation.

In the scenario A it is easy to explain the evolution of unique characters found in K (61 characters), M (46 characters) and P (5 characters) and the evolution of 59 characters shared by M and P. However, it is much more difficult to explain the evolution of characters shared by K+M (11 characters) and K+P (62 characters) because they were not found in P and M as

expected in case when these characters were inherited from the common ancestor. Therefore, to explain the pattern A we have to assume that either these 73 characters evolved in the root of the tree, but were secondary lost or these characters independently evolved in branches M and P. Thus, this scenario includes 73 homoplasies.

In the scenario B it is easy to explain the evolution of unique characters found in P (5 characters), K (61 characters) and M (46 characters) and the evolution of 11 characters shared by K and M. However, it is much more difficult to explain the evolution of characters shared by M+P (59 characters) and K+P (62 characters) because they were not found in K and M as expected in case when these characters were inherited from the common ancestor. Therefore, to explain the pattern B we have to assume that either these 121 characters evolved in the root of the tree, but were secondary lost or these characters independently evolved in branches K and M. Thus, this scenario includes 121 homoplasies.

In the scenario C it is easy to explain the evolution of unique characters found in M (46 characters), P (5 characters) and K (61 characters), and the evolution of 62 characters shared by K and P. However, it is much more difficult to explain the evolution of characters shared by K+M (11 characters) and by M+P (59 characters) because they were not found in P and K as expected in case when these characters were inherited from the common ancestor. Therefore, to explain the pattern C we have to assume that either these70 characters evolved in the root of the tree, but were secondary lost or these characters independently evolved in branches P and K. Thus, this scenario includes 70 homoplasies.

According to the hybrid speciation scenario D, 11 characters shared by K and M evolved in their common ancestor, 62 characters evolved in K before the hybridization event and were transmitted to P, 61 characters evolved in K and were not transmitted to P, 59 characters evolved in M before the hybridization event and were transmitted to P, and 46 characters evolved in M and were not transmitted to P. Finally 5 characters evolved in P after this hybrid lineage was established. Thus, the hybrid speciation scenario requires 0 (zero) homoplasies for explanation of the evolution of the AFLP markers.

Conclusion: all three scenarios of divergent speciation can not be explained without assumption about multiple homoplasies involved in evolution and are much less parsimonious than the scenario of hybrid speciation.

References

64. Hennig, W. 1965. Phylogenetic Systematics. *Ann. Rev. Entomol.* **10**, 97–116.

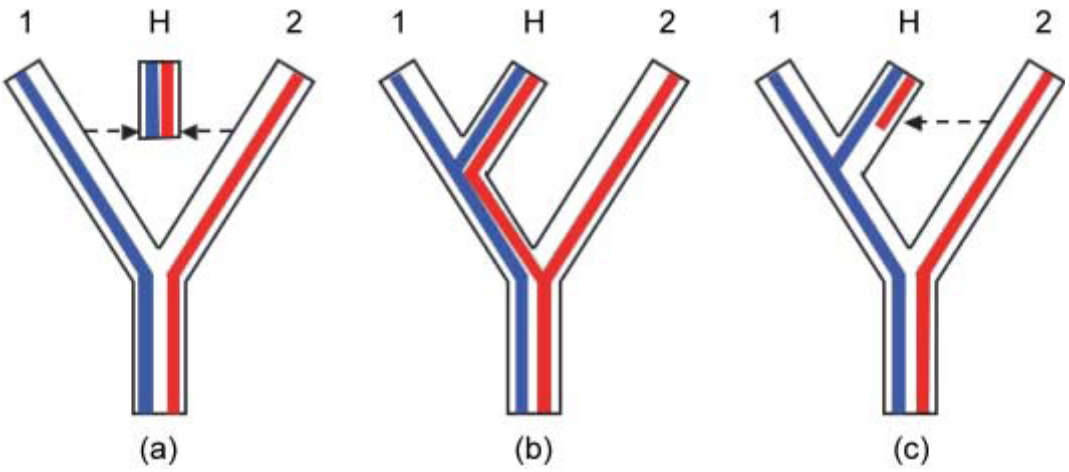
65. Krell, F.-T. 2005. A Hennigian monument on vertebrate phylogeny. *Syst. Biodiv*. **3**, 339–341.

ESM10

S10.1. An experimental approach to testing hypotheses of interspecific hybridization, shared ancestral polymorphism and introgression by using GISH

One of the main problems in studying HHS is the fact that the pattern of molecular markers predicted by HHS is difficult to distinguish from patterns predicted by inheritance of ancestral polymorphism during standard divergent speciation and by after-speciation introgression (supplementary figure S10.1).

Here we elaborated a special design of genomic *in situ* hybridization experiments that allows us, by highlighting (using labelled DNA probe) and/or suppressing (using DNA competitor) parts of genome of different origin, to make conclusions about mechanisms of genome evolution even in cases where there is no opportunity to identify the individual chromosomes. In particular, it enables to distinguish between "normal" divergent, hybrid and introgression models of genome formation by analyzing experiments in which different combinations of total molecular probes and competitors are used.



Supplementary figure S10.1. Schematic alternative speciation scenarios creating an admixed species H from species 1 and 2. (*a*) H is the product of a homoploid hybrid speciation (HHS) between species 1 and 2. (b) H is the sister species of 1, sharing ancestral polymorphisms with 1 and 2. (*c*) H is the sister species of 1, receiving subsequent adaptive introgression from 2. Scenario b implies no hybridisation, and in theory can be distinguished from the other two, although with varying degrees of difficulty depending on the genetic markers and analyses used. Scenario c is intermediate between a and b and can be particularly difficult to distinguish from the first (after [7]).

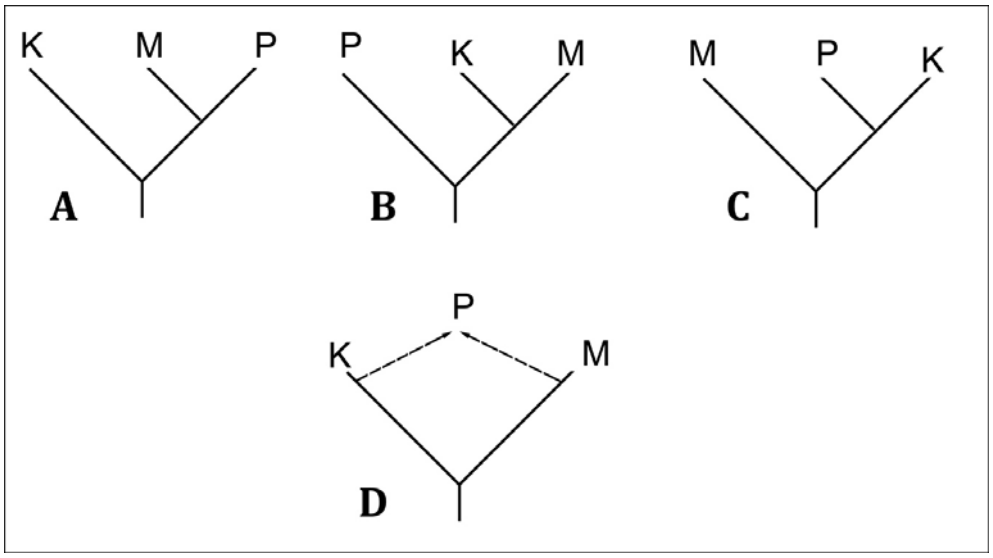
Briefly, scenario of hybrid speciation (supplementary figure S10.1*a* and supplementary figure S10.2*d*) can be differentiated from scenario of phyletic speciation (supplementary figure S10.1*b* and supplementary figure S10.2*a*-*c*), since the scenario of hybrid speciation predicts the

presence of hybridization signal in both reciprocal variants of the probe/competitor treatment, whereas phyletic evolution scenarios predict no hybridization signals or hybridization signals only in one variant of the probe/competitor treatment.

Scenario of pure hybrid speciation (Supplementary figure S10.1*a*) can be also differentiated from scenario of introgressive speciation (supplementary figure S10.1*c*) using GISH. To make this, the total labelled DNA of potential hybrid species should be used as a probe in presence of competitors derived from both parental species 1 and parental species 2 as described in the main text of the article.

S9.2. Elaborating the design of GISH experiments: distinguishing between divergent and hybrid models of speciation

Let us assume that phylogenetic relationships between K, M and P are unknown, and we are not sure of the hybrid origin of P. In this case four scenarios of evolution are possible (supplementary figure S10.2*a*-*d*), of which the first three (*a*-*c*) represent the "normal" divergent speciation while the fourth (*d*) represents hybrid speciation.



Supplementary figure S10.2. Theoretically possible phylogenetic relationships between K, M and P in cases of divergent and hybrid speciation. **(***a*-*c***)** Scenarios of divergent speciation. **(***d***)** Scenario of hybrid speciation.

Let us also assume that we analyze the chromosome preparations of P by (1) using probe of K and competitor from M, and (2) using probe of M and competitor from K (supplementary figures 10.3*a*-*d*).

Scenario A (supplementary figure S10.3*a*)

Variant 1. K=probe, M=competitor. M suppresses DNA shared by M, K and P. After suppressing, the preparation P has no molecular targets for the probe K since all the unsuppressed regions of P evolved independently of K.

Expected result: no hybridization signal will be observed.

Variant 2. K=competitor, M=probe. K suppresses DNA shared by K, M and P. However, after suppressing, the preparation P shares common repeats with the probe M that evolved after separation of the common ancestor of M+P from common ancestor of K+M+P.

Expected result: hybridization signal of M in the preparation P will be observed.

Scenario B (supplementary figure S10.3*b*)

Variant 1. K=probe, M=competitor. M suppresses DNA shared by M, K and P. After suppressing, the preparation P has no molecular targets for the probe K since all the unsuppressed regions of P evolved independently of K.

Expected result: no hybridization signal will be observed.

Variant 2. K=competitor, M=probe. K suppresses DNA shared by K, M and P. After suppressing, the preparation P has no molecular targets for the probe M since all the unsuppressed regions of P evolved independently of M.

Expected result: no hybridization signal will be observed.

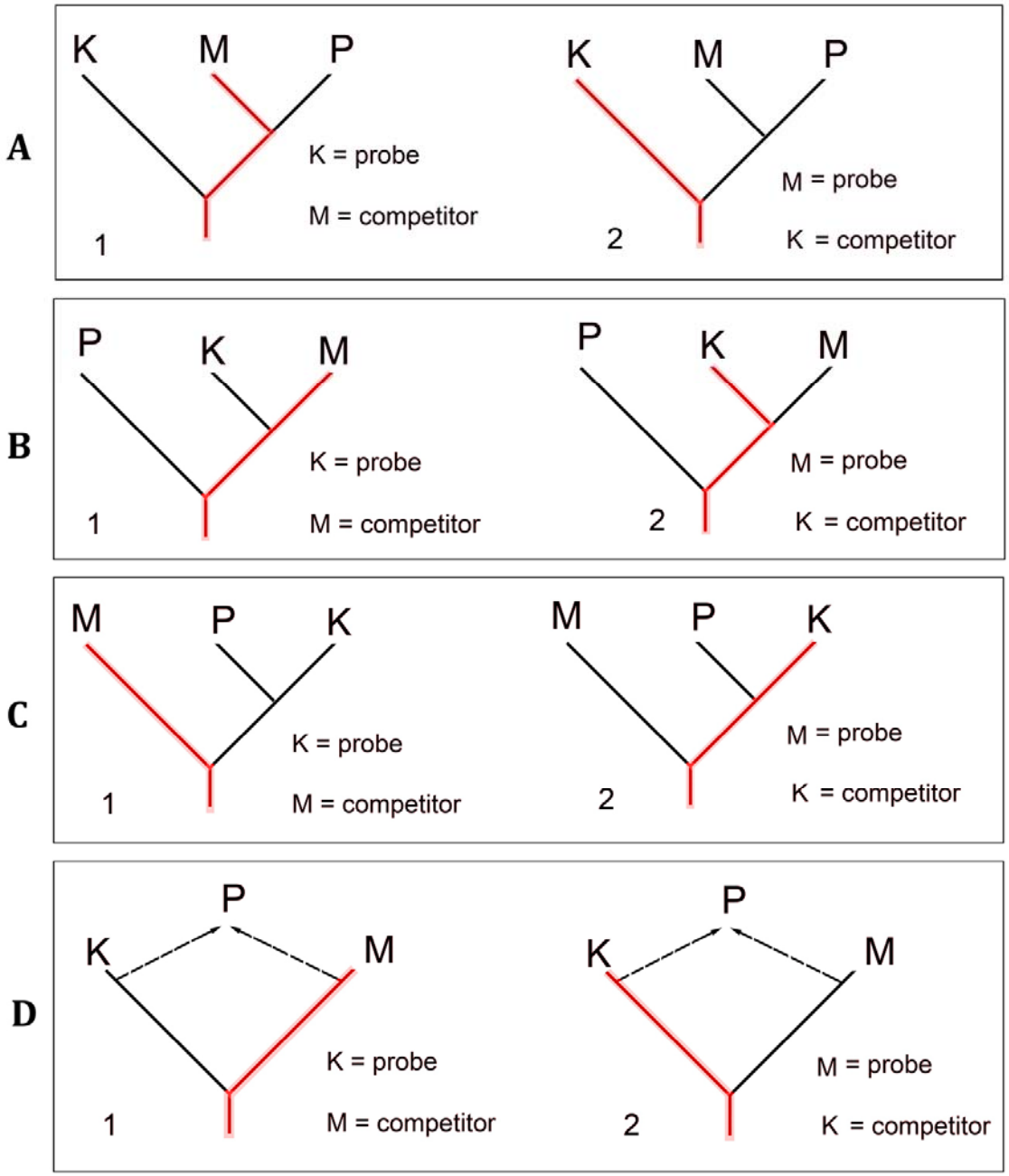
Scenario C (supplementary figure S10.3*c*)

Variant 1. K=probe, M=competitor. M suppresses DNA shared by M, K and P. However, after suppressing the preparation P shares common repeats with the probe K evolved after separation of the common ancestor of P+K from common ancestor of M+P+K.

Expected result: hybridization signal of K on the preparation P will be observed.

Variant 2. K=competitor, M=probe. K suppresses DNA shared by K, M and P. After suppressing, the preparation P has no molecular targets for the probe M since all the unsuppressed regions of P evolved independently of M.

Expected result: no hybridization signal will be observed.



Supplementary figure S10.3 Expected results of GISH experiments for the four theoretically possible scenarios of evolution. Hybridization of molecular probes (K or M) with the sequences evolved in the highlighted lineages (red) will be suppressed by competitor (M or K). **(***a-c***)** Scenarios of "normal" divergent speciation. **(***d***)** Scenario of hybrid speciation.

Scenario D (supplementary figure S10. 3*d*)

Variant 1. K=probe, M=competitor. M suppressed the regions inherited from M. After suppressing, the regions inherited from K will hybridize with the probe K.

Expected result: hybridization signal will be observed.

Variant 2. K=competitor, M=probe. K suppressed the regions inherited from K. After suppressing, the regions inherited from M will hybridize with the probe M.

Expected result: hybridization signal will be observed.

Conclusion. Scenario D (hybrid speciation) can be differentiated from scenarios A, B and C (phyletic speciation), since the scenario D predicts the presence of hybridization signal in both variants of the probe/competitor treatment, whereas scenarios A, B and C predict no hybridization signals or hybridization signals only in one variant of the probe/competitor treatment.

ESM11

GISH experiments revealed hybrid origin of P

The shared mitochondrial haplotypes, unusual colour and intermediate karyotype structure of P indicated a possible hybrid origin of P from K and M. To test this hypothesis we performed genomic *in situ* hybridization (GISH) experiments. GISH is a molecular cytogenetic technique in which labelled total DNA of parental species is hybridized to chromosome preparation of a hybrid, enabling highly effective detecting parental chromosomes and/or parental chromosome blocks in the hybrid karyotypes [54-56]. The great advantage of GISH is the fact it enables direct physical mapping multiple species-specific DNA sequences on chromosomes and detection of recombination events, representing thus the most adequate tool for analyzing hybrid karyotypes [54-56].

When the parental species are closely related, genomic probes less readily label individual parental chromosome sets in their hybrid because the number of shared DNA sequences, inherited from a common ancestor, can significantly exceeds the number of species-specific sequences evolved after separation of parental species. This difficulty can be successfully overcome by combining the labelled DNA probe of one parental species with pre-hybridization with a total blocking (=competitor) DNA obtained from the other parental species. The blocking DNA prevents *in situ* probe labeling to common sequences and thus enables parental chromosomes from more closely related species to be distinguished [54-56].

In the first series of experiments we applied the method of **self-GISH** when chromosomes and a labelled probe from the same species (K, M or P) were used, and no blocking DNA was added. In all cases we observed clear hybridization signals on all chromosomes of every species (supplementary figure S11*a*, *b* for P). We interpret this result as a positive control demonstrating that the GISH-technique was working.

In the second series of experiments we applied the method of self-GISH when chromosomes and the labelled probe from the same species (K, M or P) were used, and blocking DNA from the same species was added in a concentration gradient. In all cases we observed that in appropriate concentration the blocking DNA suppressed the hybridization signal of the labelled DNA (supplementary figure S11*c*, *d*). We interpret this result as a negative control demonstrating that the DNA blocking technique was working.

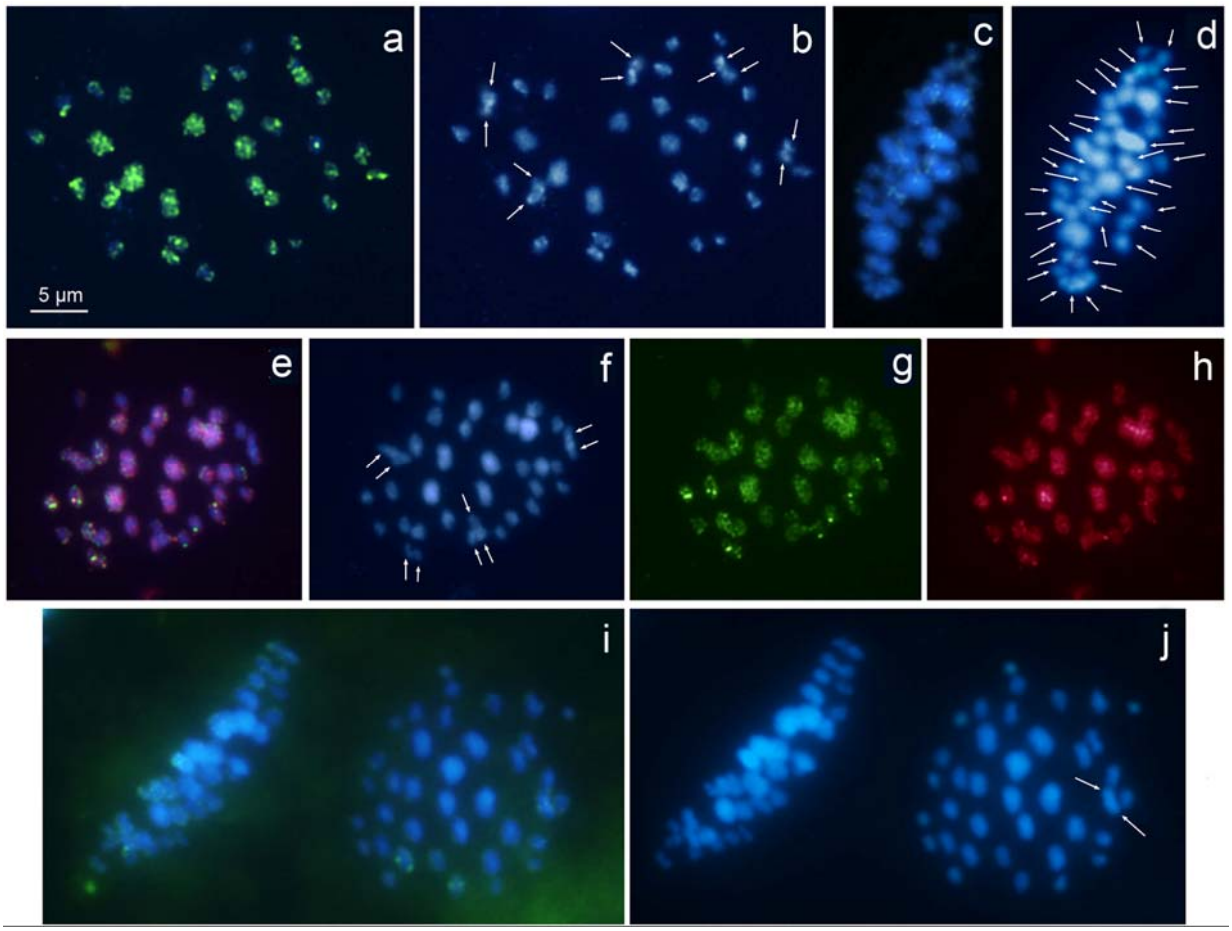


Figure S11. Self-GISH and Comparative Genomic Hybridization (CGH) in meiotic metaphase I spermatocytes of P. (*a*) Self-GISH without blocking DNA. GISH image of P-derived total genomic probe (green signals) hybridized to meiotic chromosomes of P counterstained with DAPI (blue). (*b*) DAPI image of the Fig. 11a displaying 38 bivalents of P (the overlapping bivalents are shown by arrows). (*c*) Self-GISH with blocking DNA. A probe cocktail containing P-derived total genomic DNA (green) and an excess of P-derived unlabelled total genomic DNA (competitor) hybridized to meiotic chromosomes of P counterstained with DAPI (blue). (*d*) DAPI image of the Fig. 11c displaying 38 bivalents of P (indicated by arrows). (**e**) CGH. Merged image of both M-derived (red) and K-derived (green) total genomic probes hybridized to meiotic chromosomes of P counterstained with DAPI (blue). (*f*) CGH. DAPI image of the Fig. 11e displaying 38 bivalents of P (the overlapping bivalents are shown by arrows). (*g*) CGH. Image of biotin-labelled K-derived total genomic probes hybridized to meiotic chromosomes of P. (*h*) CGH. Image of rhodamine-labelled M-derived total genomic probes hybridized to meiotic chromosomes of P. (*i*) SelfGISH with blocking DNA. A probe cocktail containing P-derived total genomic DNA (green) in combination with both K-derived and M-derived DNA competitors hybridized to meiotic chromosomes of P counterstained with DAPI (blue). Hybridization of the labelled DNA is not completely suppressed. (*j*) DAPI image of the 11ig. 4i displaying 38 bivalents of P. Two overlapping bivalents are shown by arrows.

Then we applied **Comparative Genomic Hybridization (CGH).** CGH is a variant of GISH in which the differently labelled genomic probes of both parental species compete for hybridization with chromosomes of the hybrid species. This technique enables the observation of the hybridization signals from both parents in the same preparation whereas the conventional

GISH does not. We simultaneously hybridized the differently labelled genomic probes of both K and M on chromosomes of P without adding the blocking DNA. As a result, we observed a mixture of hybridization signals on chromosomes of P. Hybridization signals of the M-derived total genomic probe (red) tended to be more common on the larger chromosomes of P located in the centre of the meiotic metaphase, and hybridization signals of the K-derived total genomic probe (green) tended to be located on the smaller chromosomes at the periphery (supplementary figure S11*e*-*h*), but the pattern was generally difficult to interpret with confidence. This result was anticipated since K and M are closely related and the no-blocking-DNA technique is therefore inapplicable for study of their genomes. Based on this result, further GISH experiments were conducted using a combination of probe and competitor DNA.

In the main series of experiments we used the technique of **conventional GISH** when the labelled DNA probe of one presumed parental species (K or M) hybridized with chromosomal preparation of P in the presence of competitor total DNA of the other parental species (M or K). We used a special design of GISH experiments that allowed us to discriminate between "normal" divergent, hybrid and introgression models of genome formation even without identification of individual chromosomes (supplementary information S10). This design is based on different predictions of hybrid and phyletic speciation modes (supplementary figures S10.1-S10.3).

According to the elaborated design, the chromosome preparations of P were used in hybridization with both K-probe/M-competitor and M-probe/K-competitor probe cocktails. In total, chromosome preparations of 20 specimens of P were analyzed in two completely independent research trials. In both trials, numerous metaphase I and metaphase II plates and diakinetic nuclei demonstrated the following two highly reproducible features. First, in all cases the strong hybridization signals appeared in both K-probe/M-competitor and M-probe/K-competitor treatments indicating a hybrid origin of the P genome. Second, the distribution of the signals was very different in these treatments. When labelled total molecular probe from M hybridized with chromosome preparations of P in the presence of the K-derived competitor, the great majority of hybridization fluorescent signals appeared on the larger chromosomes of P located in the centre of metaphase plates (figure 2*d*, *f* in the main body of the article). When labelled total molecular probe from K hybridized with chromosome preparations of P in the presence of the M-derived competitor, the hybridization signals appeared only on the small chromosomes of P located at the periphery (figure 2*e*, *g*). This result can only be explained by hybrid origin of the genome in P that inherited the majority of its small-sized chromosomes from K and the majority of its large-sized chromosomes from M, and these chromosomes in P were largely non-recombinant. This conclusion is highly consistent with the fact that the great majority of chromosomes in K are small, and the majority of chromosomes in M are large.

Since the chromosomes in P were anonymous (without specific markers, except for large or small size) and since the K and M probes were hybridized separately to different slides, we cannot exclude that few chromosomes could hybridize with both K and M probes. However, the number of such chromosomes should be low (if they were present at all) because the separation of chromosomes of these two size classes is generally easy, especially taking into account the specific architecture of lepidopteran spermatocyte metaphase I with large bivalents located in the centre and small bivalents situated at the periphery [49]. Thus, not only chromosome size, but also chromosome position can be taken into account for distinguishing chromosomes of these two size-classes.

Finally we tested the introgression model of genome formation, i.e. the hypothesis that P is not a pure hybrid species, but represents an independent, third phylogenetic lineage contaminated by interspecific gene flow from K and M through sporadic post-speciation hybridization events. To test this hypothesis, we conducted an additional series of GISH experiments in which the labelled P-derived probe hybridized to the preparations of P in presence of both K-derived and M-derived DNA competitors. If P consists of genomes of K and M only, one can expect that no hybridization signals will appear in the preparation. If P consists of not only genomes of K and M, but also includes elements of a third genome, one can expect that the hybridization signals will reveal this third genome. Analysis of the preparations revealed that the hybridization signals were present being however strongly reduced (supplementary figure S11*i*,*j*). This result is generally consistent with the explanation that P includes only chromosomes inherited from K and M. The presence of sporadic, weak hybridization signals can likely be explained by divergence of DNA repeats in P after the hybrid speciation event.