BMC Plant Biology

CrossMark

Genetic origin and composition of a natural hybrid poplar *Populus × jrtyschensis* from two distantly related species

Dechun Jiang[1†], Jianju Feng[1,2†], Miao Dong[1], Guili Wu[1], Kangshan Mao[1] and Jianquan Liu[1*]

Abstract

Background: The factors that contribute to and maintain hybrid zones between distinct species are highly variable, depending on hybrid origins, frequencies and fitness. In this study, we aimed to examine genetic origins, compositions and possible maintenance of *Populus × jrtyschensis*, an assumed natural hybrid between two distantly related species. This hybrid poplar occurs mainly on the floodplains along the river valleys between the overlapping distributions of the two putative parents.

Results: We collected 566 individuals from 45 typical populations of *P. × jrtyschensis*, *P. nigra* and *P. laurifolia*. We genotyped them based on the sequence variations of one maternally inherited chloroplast DNA (cpDNA) fragment and genetic polymorphisms at 20 SSR loci. We further sequenced eight nuclear genes for 168 individuals from 31 populations. Two groups of cpDNA haplotypes characteristic of *P. nigra* and *P. laurifolia* respectively were both recovered for *P. × jrtyschensis*. Genetic structures and coalescent tests of two sets of nuclear population genetic data suggested that *P. × jrtyschensis* originated from hybridizations between the two assumed parental species. All examined populations of *P. × jrtyschensis* comprise mainly F_1 hybrids from interspecific hybridizations between *P. nigra* and *P. laurifolia*. In the habitats of *P. × jrtyschensis*, there are lower concentrations of soil nitrogen than in the habitats occupied by the other two species.

Conclusions: Our extensive examination of the genetic composition of *P. × jrtyschensis* suggested that it is typical of F_1-dominated hybrid zones. This finding plus the low concentration of soil nitrogen in the floodplain soils support the F_1-dominated bounded hybrid superiority hypothesis of hybrid zone maintenance for this particular hybrid poplar.

Keywords: Hybrid origin, First-generation hybrids, Hybrid superiority, *Populus × jrtyschensis*

Background

Interspecific hybridization occurs frequently in plants [1–3]; hybrid swarms or hybrid zones provide a window through which to examine species cohesiveness, interspecific gene flow and hybrid fitness [4]. However, it is still hotly debated how such hybrid zones are maintained, mainly because of conflicting views about the relative role of selection versus gene flow in driving or homogenizing divergence [5]. Up to now, three types – tension zones, bounded hybrid superiority zones, and

mosaic hybrid zones – have been tentatively suggested, based on theoretical and empirical studies of how selection acts on hybrids and parent species [6, 7]. Within tension zones, hybrids are of low fitness relative to parent species and hybrid zones are restricted to a narrow area between the two parents and are mainly maintained by a balance between dispersal and selection against hybrids [8]. The bounded hybrid superiority (also called the environment-dependent) model assumes that hybrids are fitter than their parents in intermediate habitats, but less fit than parent species in their respective native habitats [9–11]. Gene flow can also be prevented if hybridization proceeds only to the F_1 stage and no further, which can occur due to apparent habitat-mediated superiority of F_1s over other hybrid classes [12]. These

* Correspondence: liujq@lzu.edu.cn
†Equal contributors
[1]State Key Laboratory of Grassland Agro-Ecosystem, School of Life Sciences, Lanzhou University, Lanzhou, Gansu, People's Republic of China
Full list of author information is available at the end of the article

Jiang et al. BMC Plant Biology (2016) 16:89

Page 2 of 12

hybrid zones probably occupy distinct habitats located in an intermediate position, where the ranges of the two parent species overlap. Finally, the mosaic hybrid zone model hypothesizes that patchy environments within the overlapping region of two parent species are highly heterogeneous [7, 13]. Therefore, hybrids comprise a mosaic of diverse genotypes that are highly variable according to their respective distributions. In such a model, both environment-independent and -dependent selections against hybrids co-exist, thus combining the hypotheses of both the tension zone model and the bounded hybrid superiority model. Whichever model applies, it is very important to know the genetic composition of such hybrid zones, with regard to genotype frequencies, before we can identify the factors that may contribute to maintaining these hybrid populations as a result of either intrinsic or extrinsic fitness.

Hybridization and gene flow between species occur extensively in the genus *Populus*, resulting in numerous natural hybrid zones [14–17]. In both Europe [15, 16, 18] and North America [14, 17, 19], the origins of numerous such hybrids have been explored. Some natural poplar hybrid zones contain a mix of F_1s, post-F_1s (F_2s) and further backcross genotypes with diverse levels of fitness [14, 17, 19–21], consistent with a combination of the hybrid tension and superiority hypotheses. Moreover, some natural poplar hybrid zones play a significant role in bridging or preventing gene flow between hybridizing species [14–22]. However, little attention has, so far, been paid to natural hybrids occurring in Asia. In this study, we aimed to examine the genetic origin, composition and possible maintenance of the hybrid between *Populus nigra* and *P. laurifolia* at numerous locations in western China. *Populus nigra*, the black poplar of sect. *Aigeiros*, is mainly found in Europe and has limited ranges in central Asia and northwest Africa [23, 24]. It is, however, a tree of social and economic importance [25]. In western China, it occurs on wet slopes beside rivers at altitudes between 400 m and 1000 m [26]. In contrast, *P. laurifolia* of sect. *Tacamahaca* occurs mainly in northern Asia, with its range extending into central Asia [27]. This species grows on the mountainous slopes of river valleys in western China; it prefers relatively dry habitats at altitudes between 400 m and 1800 m [27]. Despite their distant relationship, as revealed in all phylogenetic studies [28, 29], these two poplars co-occur in Xinjiang, western China. Both of them flower and set seed from April to May [27]. However, these two species differ from each other with respect to numerous characters from leaves to branches and flowers [27]. Both species are dioecious, with pollen dispersed by wind and seeds dispersed by wind and water [30]. They also propagate vegetatively from broken branches and cuttings [31]. Due to their overlapping distributions and flowering periods in western China, a hybrid, *P. × jrtyschensis*,
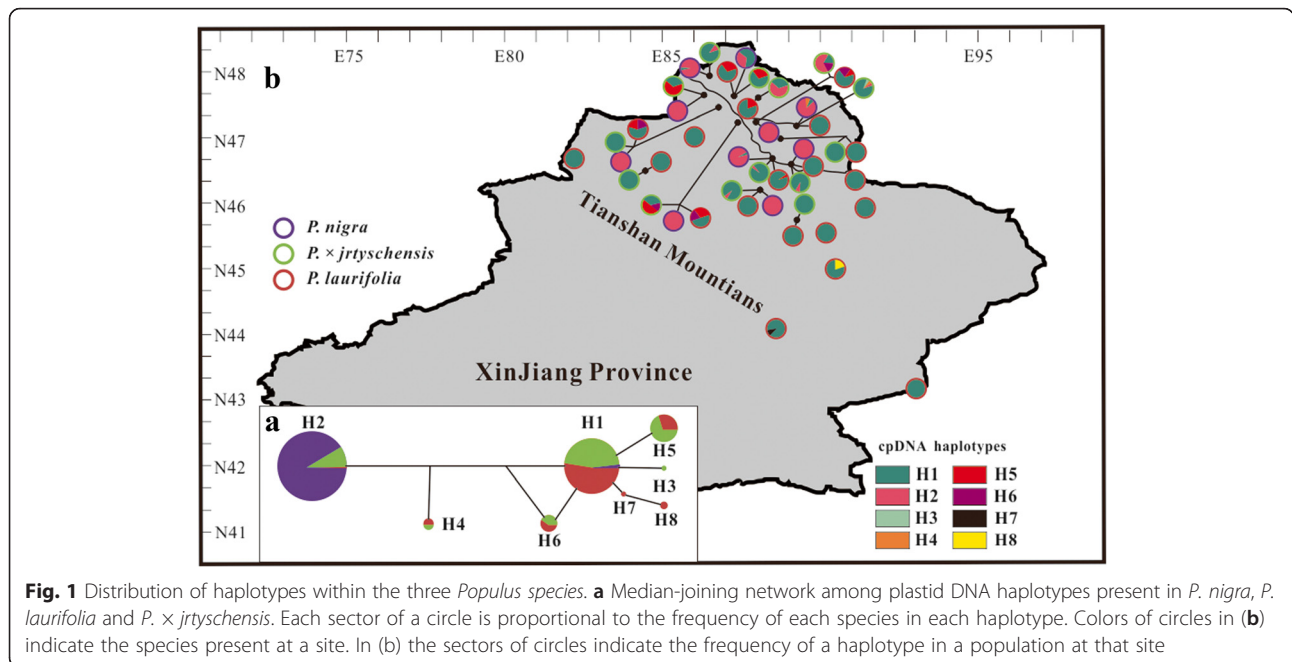
was assumed to result from crosses between these two distantly related poplars in Xinjiang [27, 32]. This hybrid and its two putative parent species are diploid with 2n = 38 [27]. It annually sets numerous seeds with unknown fertility [27]. This hybrid poplar has an intermediate morphology between *P. nigra* and *P. laurifolia*, although the overall morphology seems to be more similar to the former than the latter [27, 32]. *P. × jrtyschensis* forms pure forests in numerous locations on the floodplains along the Erqis river valley, where neither parent species is present [27, 32]. In addition, this hybrid poplar has been introduced and widely cultivated along agricultural drainage channels, by means of cuttings taken from wild populations, because of its fast growth, straight stems and the other superior characteristics compared to the putative parent species [27, 32].

In addition to the morphological evidence, genetic evidence based on sequence variations from ITS and chloroplast DNA (cpDNA) from samples of several individuals of most species found in Xinjiang has also suggested that *P. × jrtyschensis* probably originated from hybridizations between these two distantly related species [33]. We extended the example to include more natural populations of *P. × jrtyschensis* and its two putative parental species for the present study. We genotyped a total of 566 individuals from 45 populations of three taxa [see Additional file 1] based on sequence variations of the maternally inherited cpDNA and polymorphisms generated by 20 nuclear simple sequence repeat (SSR) markers. We also sequenced eight nuclear genes for 168 individuals from 31 populations. In this study, we mainly aimed to test the following hypotheses. First, *P. × jrtyschensis* originated through hybridization from two distantly related poplar species. This was investigated by examining cpDNA sequence variations and conducting coalescent analyses of genetic polymorphisms from 20 SSR and eight nuclear genes. Second, all examined populations of *P. × jrtyschensis* have the same hybrid genetic compositions probably comprised of F_1s, despite their mosaic distributions due to the relative stability in the morphology of all *P. × jrtyschensis* populations. Finally, habitat-selection contributed to the formation of these hybrid swarms and maintained them (bounded hybrid superiority hypothesis) because the floodplains where *P. × jrtyschensis* occurs is obviously poorer than the habitats of the two putative parent species. In order to confirm this, we measured and compared the soil nitrogen concentrations in typical habitats of the three taxa.

Results

Sequence variation of chloroplast DNA

Thirteen substitutions were detected at the *rbc*L gene across the 566 individuals sampled. These mutations together revealed eight haplotypes (H1-H8, [see Additional file 2]), which clustered into two major groups (Fig. 1):

Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 3 of 12

Fig. 1 Distribution of haplotypes within the three *Populus* species. **a** Median-joining network among plastid DNA haplotypes present in *P. nigra*, *P. laurifolia* and *P. × jrtyschensis*. Each sector of a circle is proportional to the frequency of each species in each haplotype. Colors of circles in (**b**) indicate the species present at a site. In (b) the sectors of circles indicate the frequency of a haplotype in a population at that site

one comprising H2 and the other consisting of H1, H3 and H5-H8. Based on the sequence variations, H4 originated from the recombination of two dominant haplotypes H1 and H2 of the two major groups. Most individuals of *P. nigra* and *P. laurifolia* were found to be fixed into a separate group of haplotypes according to species. For example, H2 was associated with most individuals of *P. nigra* but only one individual of *P. laurifolia*. In contrast, most individuals of *P. laurifolia* were H1, while this haplotype was found for only seven individuals of *P. nigra*. In addition, a few rare haplotypes (H3-H8) were found to be mainly associated with *P. laurifolia*. The individuals of *P. × jrtyschensis* that we examined were found to be represented by five haplotypes of both groups, H1, H2, H4, H5 and H6. Around 94 % of the individuals of *P. × jrtyschensis* were found to have the haplotypes mainly associated with *P. laurifolia* while 6 % were H2, which is mainly found in *P. nigra*. Genetic partitions estimated by AMOVA based on these haplotypes revealed that between-population variation was significant and accounted for 34 % of the total variation in *P. nigra*, but was not significant in *P. laurifolia* where it accounted for only 6 % of the total variation. Between-population differentiation associated with cpDNA sequence variation was significant in *P. × jrtyschensis* and accounted for 28 % of the total variation (Table 1).

Genetic diversity and structure analyses based on eight nuclear genes

Sequence variation and genetic diversity across the eight nuclear loci were both larger in *P. × jrtyschensis* than in the other two species [see Additional file 3]. Private single-nucleotide polymorphisms (SNPs) for each parent

species were recovered at each locus, with shared SNPs being more common between *P. × jrtyschensis* and *P. nigra* than between *P. × jrtyschensis* and *P. laurifolia* (Table 2). Both PCAs of samples and the NNet tree constructed for all samples suggested a hybrid origin of *P. × jrtyschensis* (Fig. 3a, b). Structure also revealed that when K was set to 2 in Structure with USEPOPINFO = 1, *P. nigra* and *P. laurifolia* individuals clustered into two separate groups, while individuals of *P. × jrtyschensis* were admixed, containing a mixture of the genomes of the two groups representing the putative parent species (Fig. 3c). Both the Pritchard et al. [34] and Evanno et al. [35] tests indicated that the most likely number of clusters for the entire data set was $K = 2$. Genetic divergence between the three taxa further indicated that *P. × jrtyschensis* was a hybrid, in that divergence between *P. × jrtyschensis* and either *P. nigra* or *P. laurifolia* was similar, while pairwise Φ_{st} values for comparisons between *P. × jrtyschensis* and either *P. nigra* or *P. laurifolia* were lower than between *P. nigra* and *P. laurifolia* (Fig. 2) [see Additional file 4]. In each taxa, the positive values for both Tajima's D and Fu & Li's D and F were estimated for half of the nuclear loci and negative for the others [see Additional file 3].

Genetic diversity and structure analyses based on SSR loci

The alleles per locus and the estimated genetic indexes for each of the three taxa were listed in [Additional file 5]. Allelic richness at each locus was higher in *P. × jrtyschensis* than *P. nigra* or *P. laurifolia* [see Additional file 6]. Both PCAs for samples from the three taxa and the NNet tree constructed for all samples based on genetic distance suggested that *P. × jrtyschensis* was located between *P. nigra*

Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 4 of 12

Table 1 Analyses of molecular variance (AMOVA) for three poplar species based on datasets for two markers

Grouping of regions	Source of variation	d.f.	SS	VC	Percent variation	Fixation index
*rbc*L haplotypes						
All samples	Among groups	2	734.907	1.925	83.32	$F_{CT} = 0.8332$
	Among populations within species	43	65.524	0.101	4.38	$F_{ST} = 0.877$
	Within populations	532	151.200	0.284	12.30	$F_{SC} = 0.2327$
P. nigra	Among populations	10	20.304	0.137	34	$F_{ST} = 0.34$
	Within populations	135	35.799	0.265	66	
P. × jrtyschensis	Among populations	13	40.935	0.183	28.87	$F_{ST} = 0.289$
	Within populations	201	90.535	0.45	71.73	
P. laurifolia	Among populations	20	4.285	0.009	6.29	$F_{ST} = 0.063$
	Within populations	196	24.867	0.127	93.71	
SSR markers						
All samples	Among species (average)	2	1535.34	2.013	28.32	$F_{CT} = 0.2832$
	Among populations within species	42	678.942	0.470	6.62	$F_{ST} = 0.3494$
	Within populations	1087	5026.09	4.624	65.06	$F_{SC} = 0.0923$
P. nigra	Among populations	10	130.972	0.376	9.22	$F_{ST} = 0.0922$
	Within populations	271	1003.85	3.704	90.78	
P. × jrtyschensis	Among populations	13	215.384	0.369	6.15	$F_{ST} = 0.0615$
	Within populations	418	2349.91	5.622	93.85	
P. laurifolia	Among populations	19	332.586	0.639	13.2	$F_{ST} = 0.132$
	Within populations	398	1672.32	4.202	86.8	
Nuclear genes						
All samples	Among species (average)	2	313.331	1.5804	53.31	$F_{CT} = -0.03546$
	Among populations within species	29	28.155	−0.04908	−1.66	$F_{ST} = 0.51659$
	Within populations	272	389.783	1.43303	48.34	$F_{SC} = 0.53314$
P. nigra	Among populations	9	82.674	−0.0513	−0.53	$F_{ST} = -0.00534$
	Within populations	82	791.608	9.6537	100.53	
P. × jrtyschensis	Among populations	13	16.811	−0.892	−10.09	$F_{ST} = -0.1009$
	Within populations	120	1167.883	9.732	110.09	
P. laurifolia	Among populations	7	28.565	−1.52901	−8.78	$F_{ST} = -0.08785$
	Within populations	70	1325.383	18.934	108.78	

d.f. degrees of freedom, *SS* sum of squares, *VC* variance components; F_{ST}, variance among populations; F_{SC}, variance within populations within groups; F_{CT}, variance among groups relative to total variance

and *P. laurifolia*, with a closer affinity with the former than the latter. The *P. × jrtyschensis* cluster was clear (Fig. 3d, e), but both the Pritchard et al. [34] and Evanno et al. [35] tests indicated that the most likely number of clusters for the entire data set was $K = 2$. When K was artificially set to 2, all individuals of *P. × jrtyschensis* were admixed, with a mixture of the genomes of the two groups representing the two parent species (Fig. 3f).

Test of the hybrid origin and hybrid composition of *P. × jrtyschensis* based on population genetic data from 20 SSRs and eight nuclear genes

We tested three alternative divergence hypotheses for the three taxa based on SSR and nuclear gene data sets separately (Fig. 5). Our ABC modeling results revealed that the hybrid origin model (Scenario 1, Fig. 5) provided a better fit for the observed data than Scenarios 2 and 3. The posterior probabilities of Scenarios 1, 2 and 3 were, respectively, 0.978, 0.004 and 0.0216 for SSRs and 0.382, 0.28 and 0.338 for the nuclear sequence dataset [see Additional file 7]. We tested hybrid composition criteria based on NewHybrids estimates suggested by Anderson and Thompson [36] using SSR and nuclear gene data sets. For SSRs, 95 % of the sampled individuals under *P. nigra* and 99 % of the sampled individuals under *P. laurifolia* were pure. In total, 84 % of the sampled individuals of *P. × jrtyschensis* were considered to be F_1 hybrids between pure *P. nigra* and *P. laurifolia*. In addition, 6 % of

Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 5 of 12

Table 2 Distribution of segregating sites at nuclear loci in pairwise comparisons of taxa: *P. nigra*, *P. × jrtyschensis* and *P. laurifolia*

Gene	*P. nigra* vs. *P. × jrtyschensis*				*P. × jrtyschensis* vs *P. laurifolia*				*P. nigra* vs *P. laurifolia*			
	S_n	S_j	S_s	S_f	S_l	S_j	S_s	S_f	S_n	S_l	S_s	S_f
Dehy	11	14	10	0	3	14	3	0	11	3	1	0
Phyto A	2	3	2	0	0	3	0	0	2	0	0	0
Phyto B	3	4	2	0	1	4	1	0	3	1	0	0
PAL	20	27	20	0	16	27	16	0	20	16	16	2
AREB1	35	39	33	0	34	39	33	0	35	34	31	0
ERD7	6	8	6	0	5	8	5	0	6	5	4	0
EIN3	30	32	30	0	28	32	28	0	30	28	26	0
LTCOR11	5	7	5	0	2	7	2	0	5	2	1	0

S_n, S_j and S_l are the number of polymorphic sites unique to *P. nigra*, *P. × jrtyschensis* and *P. laurifolia*, respectively, in each comparison; S_s is the number of sites with shared alleles between the two taxa; and S_f is the number of sites with fixed alleles in either taxa

individuals are backcrosses with one of the parents, while it is difficult to ascribe the remaining individuals (Fig. 4). Similarly, based on sequence variations of nuclear genes, 90 and 100 % of the sampled individuals under *P. nigra* or *P. laurifolia* were found to be pure. In addition, 87 % of the sampled individuals of *P. × jrtyschensis* were considered to be F_1 hybrids while 9 % of them seems to be backcrosses with one of the parents and the remaining individuals were difficult to ascribe. Only two individuals from one population were found to have the same marked polymorphisms at all 20 SSR loci, suggesting that they derived from the same clone. No single clone was found in any two different populations.

Based on SSR data sets, gene flow (N_em) was estimated to be greater from *P. laurifolia* and *P. nigra* (0.5952) than in the opposite direction (0.2218). Gene flow occurred more frequently between *P. × jrtyschensis* and the two parent species. More gene flow occurred from *P. laurifolia* to

P. × jrtyschensis (2.91) than in the reverse direction (0.8644) while less was detected from *P. nigra* (0.8944) to *P. × jrtyschensis* than in the reverse direction (3.1402). The same trend was observed based on nuclear genes: gene flow was estimated to be 0.1094, 0.005 and 0.111 separately from *P. laurifolia* to *P. nigra*, from *P. nigra* to *P. × jrtyschensis* and from *P. × jrtyschensis* to *P. laurifolia*, respectively, and in the opposite direction it was estimated to be 0.0111, 0.2044 and 0.2283. In all directions, rates of gene flow estimated for the SSR data set were greater than those based on nuclear gene sequence data (Fig. 6).

Soil nitrogen analyses of typical habitats for three taxa

Total soil nitrogen concentration of typical habitats of *P. × jrtyschensis* differed from those of the two parent species. The typical habitats of *P. × jrtyschensis* had lower nitrogen concentrations at depths of 0–20 cm, 20–40 cm and 40–70 cm than the habitats of the two parent species (Fig. 7). In addition, we found that soil nitrogen concentrations were significantly different between *P. × jrtyschensis* habitats and the habitats of the two parent species, with higher probabilities for the greater depths [see Additional files 8, 9 and 10].

Discussion

In this study, we used 20 SSR markers, eight nuclear gene markers and cpDNA sequence variations to genotype 566 individuals from 45 populations of *P. × jrtyschensis*, *P. nigra* and *P. laurifolia*. In addition to the intermediate morphology of the hybrid compared to the two putative parents [27, 32], our genetic results provided further support for the hypothesis that *P. × jrtyschensis* originated from hybridizations between the distantly related species *P. nigra* and *P. laurifolia*. Our reasons for this conclusion are as follows. First, the detected alleles for each individual of *P. × jrtyschensis* were admixed with the clusters specific to the putative parent species. That the species-specific alleles co-occurred in one taxa undoubtedly suggested its hybrid origin [34]. This scenario has been confirmed in
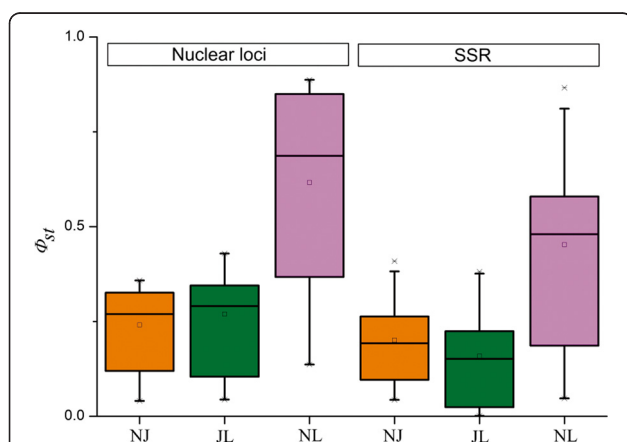

Fig. 2 Box plot of genetic differentiation (Φst) between each of the three species pairs based on nuclear and SSR data sets. NJ, JL and NL represent the Φ_{st} value between *P. nigra* and *P. × jrtyschensis*, *P. × jrtyschensis* and *P. laurifolia*, *P. nigra* and *P. laurifolia*. The Φ_{st} of each locus was estimated individually by AMOVA. Divergence between *P. × jrtyschensis* and parent species is lower than that between parent species as expected for a hybrid species
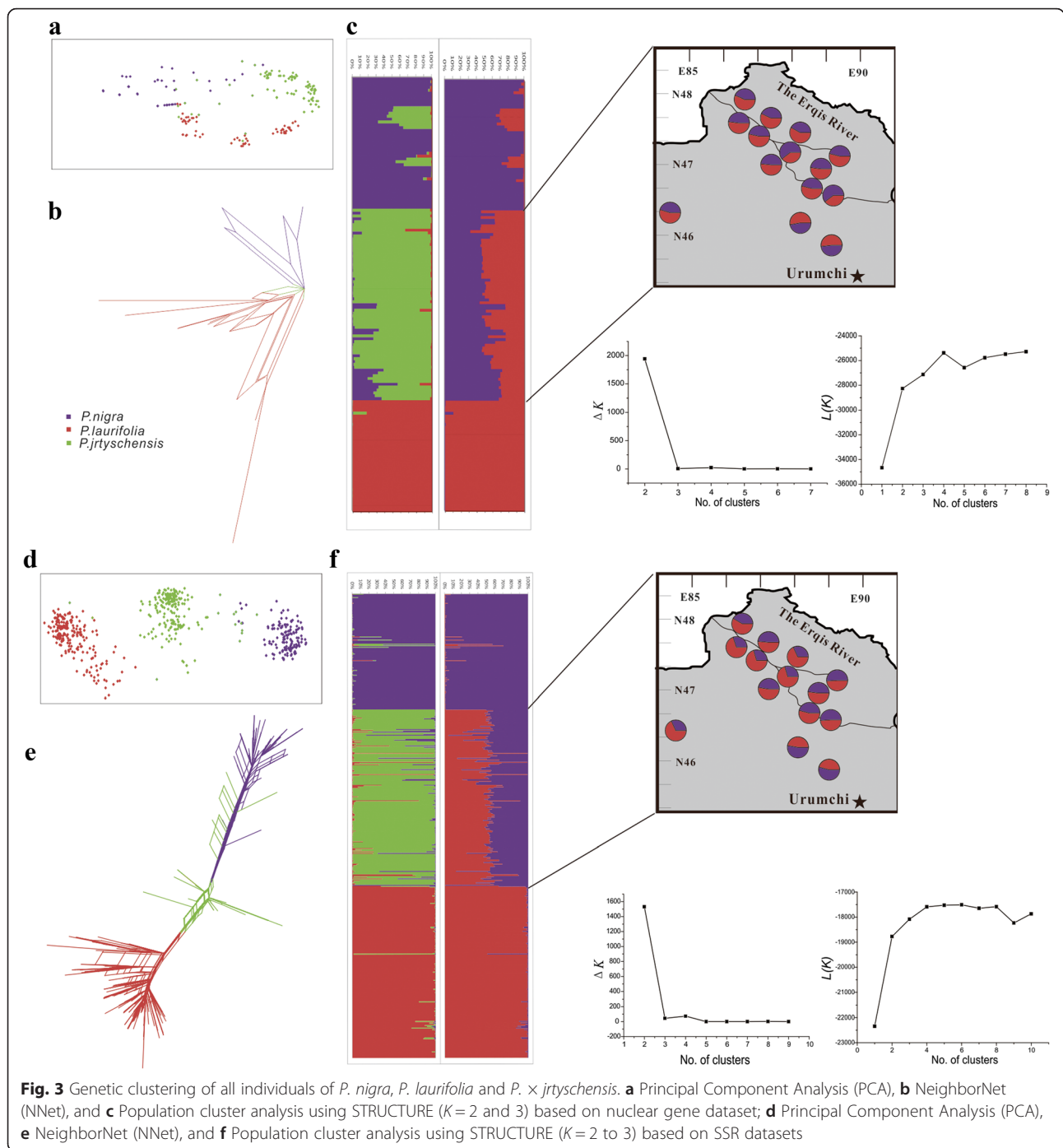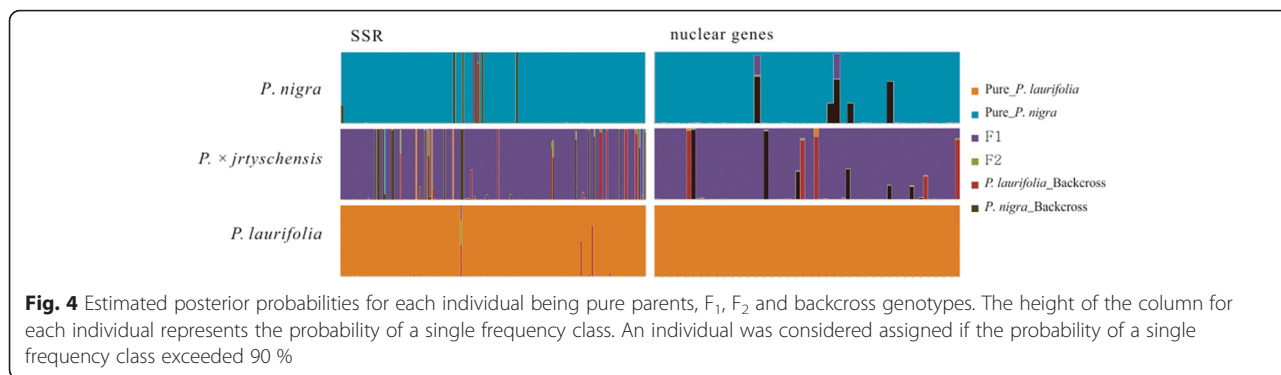
Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 6 of 12

Fig. 3 Genetic clustering of all individuals of *P. nigra*, *P. laurifolia* and *P.* × *jrtyschensis*. **a** Principal Component Analysis (PCA), **b** NeighborNet (NNet), and **c** Population cluster analysis using STRUCTURE ($K = 2$ and 3) based on nuclear gene dataset; **d** Principal Component Analysis (PCA), **e** NeighborNet (NNet), and **f** Population cluster analysis using STRUCTURE ($K = 2$ to 3) based on SSR datasets

some case study of hybrid taxa [12]. Second, ABC analyses supported the hybrid origin hypothesis for *P.* × *jrtyschensis* while the alternative hypotheses suggesting divergences from one of the two parent species were rejected (Fig. 5). Finally, two distinct cpDNA lineages were recovered for *P. nigra* and *P. laurifolia* respectively while both of them co-occurred in *P.* × *jrtyschensis*. Two divergent maternal lineages from putative parents have also been reported for other hybrid taxa [2, 3]. These lines of evidence together

supported the hypothesis that *P.* × *jrtyschensis* originated from hybridizations between *P. nigra* and *P. laurifolia*.
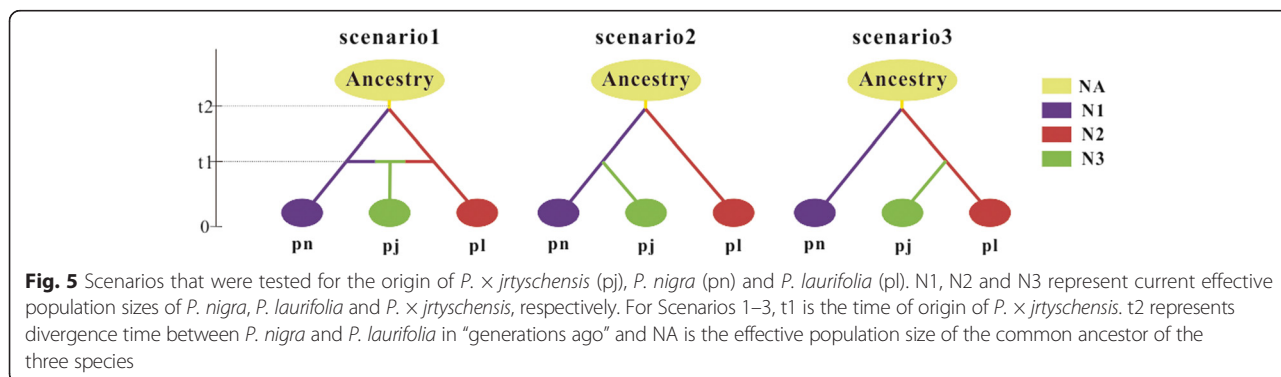
Further, we found that most of the populations of *P.* × *jrtyschensis* that we examined comprised F_1 hybrids with a few backcrosses with each of the two parent species, although clonal reproduction did occur in some of them. These findings did not support the other two original hypotheses regarding the intermediate but stable morphology of *P.* × *jrtyschensis*, namely that they either derived
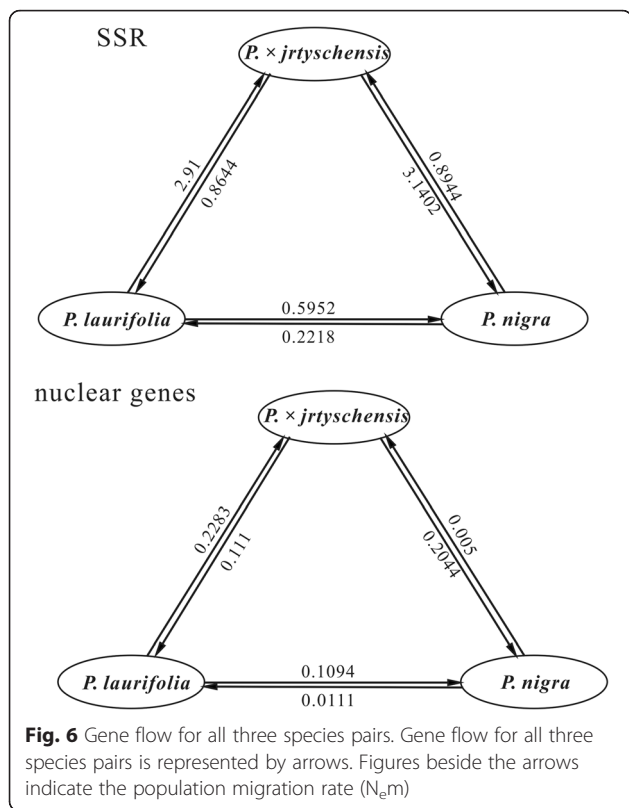
Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 7 of 12

Fig. 4 Estimated posterior probabilities for each individual being pure parents, F_1, F_2 and backcross genotypes. The height of the column for each individual represents the probability of a single frequency class. An individual was considered assigned if the probability of a single frequency class exceeded 90 %

from a few clonal lineages or had developed into a stable homoploid hybrid species. However, in a typical hybrid zone, F_1s usually comprise a very small number of the individuals present [37, 38]. Relatively few hybrid zones have been reported to be dominated by F_1s; those that were known include *Encelia × laciniata* [39], the hybrid zone between Black Oaks [40], *Rhododendron × sochadzeae* [12] and *Rhododendron agastum* [41]. A predominance of F_1s has rarely been found in hybrid swarms between other *Populus* species and most hybrid swarms contain F_1s, F_2s as well as backcrosses [14, 17, 19–21]. In a previous study [14], only F_1s were detected between *P. deltoides* and *P. nigra*, possibly due to their distant relationship and strong reproductive isolation. According to our field observations, *P. × jrtyschensis* produced numerous seeds. However, it remains unknown whether these seeds germinate. We also failed to find young seedlings from the habitat of *P. × jrtyschensis*, which seems to support the conclusion that the populations of *P. × jrtyschensis* mainly comprise F_1s. Because we did detect backcross hybrids (although fewer individuals) with both *P. nigra* and *P. laurifolia*, pollen-stigma incompatibility is unlikely to account for the general absence of the post-F_1s in most of the populations of *P. × jrtyschensis* that we examined. However, introgressions between *P. nigra* and *P. laurifolia* are relatively small according to our estimations based on the nuclear dataset (Fig. 6) despite the fact that these F_1s
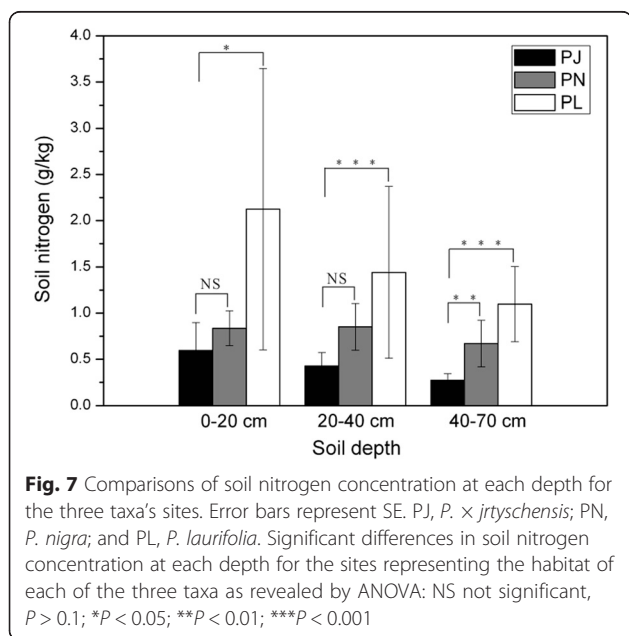
might have resulted from the repeated hybridizations between two parental species.

The presence of these mosaic hybrid populations consisting mainly of F_1s suggests two alternative origins: a recent contact between two parental species only one generation ago without enough time for post-F_1 derivatives to have been produced or that these F_1s may exclude other genotypes from the hybrid habitats [12, 37]. Numerous individuals of each examined population are at least 50 years old according to rough estimates based on their large stems compared with other poplars encountered during our field surveys. Although accurate data on flowering age of *P. × jrtyschensis* are not available, this should be similar to other poplars, i.e. between 10 and 30 years [27]. Therefore, most genets of each population should have existed long enough for post-F_1 progeny to have been produced. Thus, it appears that the *P. × jrtyschensis* populations comprise stable and long-lived hybrid zones dominated by F_1s, and other genotypes were excluded because of the habitat selection. The distributional preferences of *P. × jrtyschensis* and the two parent species also support this habitat-selection suggestion. At a local scale, *P. × jrtyschensis* is parapatric, rather than strictly sympatric to the two parent species. One of the parent species, *P. nigra*, was found on wet slopes adjacent to rivers, whilst the other, *P. laurifolia*, was found on dry mountainous slopes; in contrast,

Fig. 5 Scenarios that were tested for the origin of *P. × jrtyschensis* (pj), *P. nigra* (pn) and *P. laurifolia* (pl). N1, N2 and N3 represent current effective population sizes of *P. nigra*, *P. laurifolia* and *P. × jrtyschensis*, respectively. For Scenarios 1–3, t1 is the time of origin of *P. × jrtyschensis*. t2 represents divergence time between *P. nigra* and *P. laurifolia* in "generations ago" and NA is the effective population size of the common ancestor of the three species

Jiang et al. BMC Plant Biology (2016) 16:89

Page 8 of 12

Fig. 6 Gene flow for all three species pairs. Gene flow for all three species pairs is represented by arrows. Figures beside the arrows indicate the population migration rate (N_em)

P. × jrtyschensis occurs exclusively on the floodplains. Three examined sites with *P. × jrtyschensis* were found to be nutrient-poor with low concentrations of the total soil nitrogen, especially in the deeper layers (Fig. 7). Such differentiations of the habitat preferences have also been noted between some hybrid taxa and their respective

Fig. 7 Comparisons of soil nitrogen concentration at each depth for the three taxa's sites. Error bars represent SE. PJ, *P. × jrtyschensis*; PN, *P. nigra*; and PL, *P. laurifolia*. Significant differences in soil nitrogen concentration at each depth for the sites representing the habitat of each of the three taxa as revealed by ANOVA: NS not significant, $P > 0.1$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$

parental species for other plant genera [12, 39, 41]. The habitat-mediated selection may have prevented other genotypes (parents, BCs and F_2s) from germination and surviving in the floodplains occupied by *P. × jrtyschensis*. In addition, new and recent hybridizations between two parental species may have continuously produced more F_1s to repopulate the *P. × jrtyschensis* hybrid zones. It is highly likely that habitat-mediated selection as well as repeated productions of the F_1s between two parental species have together maintained the unique F_1 hybrid zones detected here.

Although direct comparisons of fitness between F_1s and F_2s or further backcrosses with either parent are rarely undertaken [6], a higher fitness for F_1s is theoretically likely. Complete gene sets from both parents are present in F_1s, and heterosis and hybrid vigor undoubtedly persist without hybrid breakdown [42, 43]. All beneficial traits conferred through the co-adapted gene complexes from two parents can be passed intact to the F_1 generation, but not to post-F_1s because such gene complexes are likely to be broken down. Therefore, if some of these co-adapted gene complexes confer a benefit to F_1s through heterosis when occupying new niches, then these effects will be reduced in post-F_1s due to the lower proportion of heterozygous loci, reflecting post-mating reproductive isolation between highly divergent species. However, increased fitness in the post-F_1s could derive from transgressive segregations, which give rise to beneficial traits that do not exist in the parent species, in homoploid hybrid neospecies or in plants developing into independent lineages [42, 43]. Theoretically, some post-F_1s are likely to develop superior traits over F_1s to occupy novel or arid habitats in places that do not favor F_1s, but which neither of the parents are adapted to. This may be true for *P. × jrtyschensis* although the predominance of F_1s in the patchy habitat prevents further segregations. In addition, the backcross frequencies observed here are extremely low, although we could not exclude the possibility that this was the result of widespread and strong genomic incompatibility between these highly divergent species. It is also likely that further backcross hybridizations were excluded by unfavorable epistatic combinations that led to unfit progeny. All these hypotheses and those suggesting higher fitness of the F_1s than F_2s, BCs and parents need further artificially controlled tests especially in the soils with the limited nitrogen concentration, as have recently been undertaken for spruce hybrids [44], before definitive conclusions can be drawn.

Conclusions

Our results suggest that *P. × jrtyschensis* is typical of F_1-dominated hybrid zones between the distantly related species *P. nigra* and *P. laurifolia*. Habitat-mediated selection due to F_1 superiority as well as continuous production of

Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 9 of 12

the more F$_1$s due to the repeated hybridizations between two parental specie are likely to have maintained these hybrid populations. Therefore, the formation of *P. × jrtyschensis* hybrid zones is largely consistent with the environment-dependent bounded hybrid superiority hypothesis. In addition, because of the absence of a basic difference in the genetic composition between the populations of *P. × jrtyschensis* examined, individuals for cultivation of this hybrid poplar can be obtained from vegetative cuttings from any natural population.

Methods

Ethics statement
All leave samples employed in this study were collected from tree species that are not endangered, and these trees grow in public area where no permission for collection of leaves is needed in China. All soil samples employed in this study were collected from public area where no permission is needed in China.

Sampling and sequencing
Leaves of 566 samples were collected from 45 populations of *Populus × jrtyschensis*, *P. nigra*, and *P. laurifolia* in Xinjiang, western China [see Additional files 1 and 11]. These populations cover the distributional ranges of *P. nigra* and *P. laurifolia* in Xinjiang, within which *Populus × jrtyschensis* occurs. Trees from each population (or location) were randomly sampled and an effort was made to avoid sampling closely related individuals or clones. Fresh leaves were dried and stored in silica gel, and the latitude, longitude and altitude of each collection site were recorded using an eTrex GIS unit (Garmin, Taiwan). We extracted the total DNA using the modified hexadecetyltrimethyl ammonium bromide (CTAB) procedure [45, 46]. Following DNA extraction, a total of nine DNA fragments were amplified and sequenced. These sequences included one chloroplast gene *rbc*L (for 566 individuals) and eight nuclear genes (*Dehy*, *PhytoA*, *PhytoB*, *PAL*, *AREB1*, *ERD7*, *EIN3* and *LTCOR11*) (for 168 individuals from 31 populations) [see Additional file 12]. The nuclear genes were selected and primers were designed from the genome sequences of two poplars (*Populus euphratica* Oliv. and *P. trichocarpa* Torr.) [47]. Sequences were edited and aligned manually using MEGA5 [48]. All newly obtained sequences for each taxon have been deposited in GenBank. All polymorphic and heterozygous sites were visually confirmed and separated. We further examined genetic polymorphisms of all 566 samples using 20 pairs of nuclear simple sequence repeat (SSR) primers reported before [see Additional file 13] [47, 49],

Population genetic analyses
We determined basic population genetic parameters for the eight nuclear genes using DnaSP, version 5.0 [50], after excluding insertions/deletions (indels). We estimated the number of segregating sites (S), Watterson's parameter (θ_w) [51], nucleotide diversity (π, [52]) and the minimum number of recombinant events (R$_m$, [53]). Haplotypes were investigated by estimating haplotype number (K) and diversity (H$_d$) for each gene based on the number of segregating sites [54, 55]. We tested the neutral evolution of loci using diverse statistics, including Tajima's D statistic [56], Fu and Li's D*and F* [57]. To quantify the extent of genetic divergence between species, we calculated the fixation index Φ_{st} [58], based on population genetic data for the eight nuclear loci and the 20 SSRs using ARLEQUIN version 3.0 [59], with significance determined by permutation tests involving 10,000 resamples. ARLEQUIN v.3.0 [59] was also used to quantify hierarchical genetic divergence between and within species using an AMOVA analysis based on nuclear data for the eight nuclear loci and the 20 SSRs; significance was assessed using the permutation test in the program with 1000 permutations. The NETWORK program [60] was used to construct a network of relationships between haplotypes identified for each nuclear locus and also to construct a network of cpDNA haplotypes based on sequence variation across *rbc*L fragments. The default settings were used for all other parameters.

The Bayesian model-based clustering method in STRUCTURE version 2.3.2 [35, 61] was used to examine genetic clustering of the nuclear data. In the analysis of nuclear sequence variation, only individuals ($N = 155$) with sequences for all eight loci and that were satisfactorily phased were included, while the analysis of SSR genotypes included all individuals ($N = 566$). To assign individuals to genetic groups (K), 10 replicate runs were conducted for each value of K, ranging from 1 to 10. The admixture model with correlated allele frequencies was used for each run with no prior placed on population origin. Each run included a burn-in of 500,000 followed by 2,000,000 Monte Carlo Markov chain (MCMC) iterations. The most likely number of clusters was estimated using the original method from Pritchard et al. [34] and also the ΔK statistic of Evanno et al. [35]. Graphics were produced using Origin version 8.

To detect genetic groupings further, principal component analysis (PCA) was also conducted separately on the nuclear gene and SSR data sets (using GenAlEx 6.5 [62]). We also used the Neighbor-Net algorithm (NNet) [63] within SPLITSTREE version 4.13.1 [64] to construct the phylogenetic relationships between individuals based on 11 of the nuclear genes and the SSR data set. NeighborNet networks were used to provide more detailed visualization of any potential conflicts among the analyzed genotypes. These conflicts can be the result of evolutionary events such as hybridization, polyploidization and recombination [65, 66]. The genetic

Jiang et al. BMC Plant Biology (2016) 16:89

Page 10 of 12

distances based on the SSR data set were measured by GenAlEx 6.5 [62].

Test of the hybrid origin and hybrid composition of P. × jrtyschensis

Three alternative divergence and speciation histories hypothesized for the three taxa were summarized in Fig. 5. We used population genetic data obtained from eight nuclear gene sequences and 20 SSR markers to test which of these three models provided the best fit for the data using Approximate Bayesian Computation (ABC) analysis in DIYABC, version 2.0.4 [67, 68]. We set the order of evolutionary relationships between P. × jrtyschensis, P. nigra, and P. laurifolia using uniform population-size parameters and timing parameters for dating divergence and hybridization. In the hybridization model, P. × jrtyschensis originated from a hybrid population between the other two species. In the other two scenarios, P. × jrtyschensis diverged from a common ancestor with one of the other two species. To select the model that best explained the genetic polymorphism observed in the three varieties, 1,000,000 multilocus genetic data sets were simulated for each scenario. We used the 1 % of the simulated data sets closest to the observed data to estimate the relative posterior probability [with 95 % confidence intervals (CIs)] for each scenario via logistic regression and posterior parameter distributions according to the most likely scenario [67, 68]. Mutation rates were assumed to be between 10^{-4} and 10^{-3} substitutions/site/year [69].

In addition, we checked whether each population of P. × jrtyschensis comprised an F_1 generation or further backcrosses with each parent species, using NewHybrids Version 1.0 [36] to estimate posterior probabilities for each individual being pure parental, F_1, F_2 or backcrossed genotypes based on the SSR and nuclear gene data sets. An individual was considered assigned if the probability of a single frequency class exceeded 90 %. We assumed that the sampled individuals originated from the same clone if they shared the same genetic polymorphisms at the 20 loci examined. We used Genclone 2.0 to detect clone individuals across all 45 populations.

Finally, we used the coalescent-based program IMa2 [70, 71] to estimate gene flow between the three taxa using the SSR and nuclear genes data sets. The mutation rate was assumed to be 10^{-4} substitutions/site/year [69] and was input as a point estimate. Average generation time was set to 15 years based on previous estimates for poplar trees [72].

Soil nitrogen analysis

Soil samples were randomly collected by taking 5-cm-diameter soil cores from 0 to 20, 20 to 40 and 40 to 70 cm depths from nine typical sites for the three taxa (three different sites for each taxon, [see Additional file 14]). All samples were dried at 105 °C to constant weight and passed through a 1 mm sieve prior to nitrogen analysis. Total nitrogen of each soil sample was determined using a Nitrogen Analyzer System (KJELTEC 2300 AUTO SYSTEM II). All statistical analyses were carried out in the SPSS statistical software package (SPSS Inc. Chicago, IL, USA). Graphics were produced using Origin version 8.

Consent to Publish
Not applicable.

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and its additional files. The nuclear genes sequences datasets are available in the Genebank repository Accessions no: KT626975-KT629820 [see Additional file 15] [http://www.ncbi.nlm.nih.gov/genbank/].

Additional files

> **Additional file 1:** The geographical distribution of populations. (PDF 48 kb)
>
> **Additional file 2:** Variable sites of the aligned sequences of chloroplast DNA fragments in eight haplotypes of P. nigra, P. laurifolia and P. × jrtyschensis. (PDF 29 kb)
>
> **Additional file 3:** Nucleotide variation, haplotype diversity and neutrality tests at eight nuclear loci in (a) P. nigra, (b) P. × jrtyschensis, (c) P. laurifolia. (XLS 36 kb)
>
> **Additional file 4:** Φ_{st} values for each taxa pair at eight nuclear loci and for the SSR data set. (PDF 147 kb)
>
> **Additional file 5:** Estimates of genetic diversity (per locus) for each species and their hybrid based on 20 microsatellite loci. (XLS 36 kb)
>
> **Additional file 6:** Characteristics of 20 microsatellite loci scored in the 566 individuals. (PDF 17 kb)
>
> **Additional file 7:** Description of all scenarios used in the Approximate Bayesian Computation analysis in DIYABC v2.0.4 to test the hybrid origin. (PDF 164 kb)
>
> **Additional file 8:** Summary results of one-way ANOVAs to determine the significant differences in soil nitrogen concentration at a depth of 0–20 cm for three taxa. (PDF 99 kb)
>
> **Additional file 9:** Summary results of one-way ANOVAs to determine the significant differences in soil nitrogen concentration at a depth of 20–40 cm for three taxa. (PDF 98 kb)
>
> **Additional file 10:** Summary results of one-way ANOVAs to determine the significant differences in soil nitrogen concentration at a depth of 40–70 cm for three taxa. (PDF 173 kb)
>
> **Additional file 11:** Sample information for each population. (XLS 30 kb)
>
> **Additional file 12:** The genes and primers used in this study. (PDF 162 kb)
>
> **Additional file 13:** SSR Primers used in this study. (PDF 99 kb)
>
> **Additional file 14:** The three different sites for each taxon used for the soil nitrogen analysis. (PDF 152 kb)
>
> **Additional file 15:** GenBank Accession numbers for all newly obtained sequences for the three taxa included in this study. (PDF 17 kb)

Abbreviations
cpDNA: chloroplast DNA; BCs: backcrosses; SMM: stepwise mutation model; IAM: infinite allele model; NNet: NeighborNet; MCMC: Monte Carlo Markov

Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 11 of 12

Chain; PCA: principal component analysis; ABC: Approximate Bayesian Computation.

Author details
[1]State Key Laboratory of Grassland Agro-Ecosystem, School of Life Sciences, Lanzhou University, Lanzhou, Gansu, People's Republic of China. [2]College of Plant Sciences, Xinjiang Production & Construction Corps Key Laboratory of Protection and Utilization of Biological Resources in Tarim Basin, Tarimu University, Alar, Xinjiang, People's Republic of China.

References

1. Stace CA. Hybridization and the plant species. In: Urbanska KM, editor. Different Pattern Higher Plants. New York: Academic; 1987. p. 115–27.
2. Rieseberg LH, Carney SE. Plant hybridization. New Phytol. 1998;140:599–624.
3. Mallet J. Hybridization as an invasion of the genome. Trends Ecol Evol. 2005; 20:229–37.
4. Hird S, Reid N, Demboski J, Sullivan J. Introgression at differentially aged hybrid zones in red-tailed chipmunks. Genetics. 2010;138:869–83.
5. Barton NH. The role of hybridization in evolution. Mol Ecol. 2001;10:551–68.
6. Arnold ML, Hodges SA. Are natural hybrids fit or unfit relative to their parents? Trends Ecol Evol. 1995;10:67–71.
7. Abbott RJ, Brennan AC. Altitudinal gradients, plant hybrid zones and evolutionary novelty. Phil Trans Roy Soci B: Biol Sci. 2014;369:20130346.
8. Barton NH. The dynamics of hybrid zones. Heredity. 1979;43:341–59.
9. Harrison RG. Hybrid zones: windows on evolutionary process. Oxf Surv Evol Biol. 1990;7:69–128.
10. Miglia KJ, Mcarthur ED, Moore WS, Wang H, Graham JH, Freeman DC. Nine-year reciprocal transplant experiment in the gardens of the basin and mountain big sagebrush (*Artemisia tridentata: Asteraceae*) hybrid zone of Salt Creek Canyon: the importance of multiple-year tracking of fitness. Biol J Linn Soc. 2005;86:213–25.
11. Goulson D. Evaluating the role of ecological isolation in maintaining the species boundary between *Silene dioica* and *S. latifolia*. Plant Ecol. 2009;205: 201–11.
12. Milne RI, Terzioglu S, Abbott RJ. A hybrid zone dominated by fertile F1s: maintenance of species barriers in *Rhododendron*. Mol Ecol. 2003;12:2719–29.
13. Arnold ML, Bulger MR, Burke JM, Hempel AL, Williams JH. Natural hybridization: how low can you go and still be important? Ecology. 1999;80: 371–81.
14. Thompson SL, Lamothe M, Meirmans PG, Périnet P, Isabel N. Repeated unidirectional introgression towards *Populus balsamifera* in contact zones of exotic and native poplars. Mol Ecol. 2010;19:132–45.
15. Lexer C, Fay MF, Joseph JA, Nica M, Heinze B. Barrier to gene flow between two ecologically divergent populus species, *P. alba* (white poplar) and *P. tremula* (European aspen): the role of ecology and life history in gene introgression. Mol Ecol. 2005;14:1045–57.
16. Van Loo M, Joseph JA, Heinze B, Fay MF, Lexer C. Clonality and spatial genetic structure in *Populus × canescens* and its sympatric backcross parent *P. alba* in a central European hybrid zone. New Phytol. 2008;177:506–16.
17. Floate K. Extent and patterns of hybridization among the three species of *Populus* that constitute the riparian forest of southern Alberta. Can J Bot. 2004;82:253–64.
18. Arens P, Coops H, Jansen J, Vosman B. Molecular genetic analysis of black poplar (*Populus nigra* L.) along dutch rivers. Mol Ecol. 1998;7:11–8.
19. Martinsen GD, Whitham TG, Turek RJ, Keim P. Hybrid populations selectively filter gene introgression between species. Evolution. 2009;55:1325–35.
20. Hamzeh M, Sawchyn C, Perinet P, Dayanandan S. Asymmetrical natural hybridization between *Populus deltoides* and *P. balsamifera* (Salicaceae). Can J Bot. 2007;85:1227–32.
21. Keim P, Paige KN, Whitham TG, Lark KG. Genetic analysis of an interspecific hybrid swarm of *Populus*: occurrence of unidirectional introgression. Genetics. 1989;123:557–65.
22. Vanden BA, Storme V, Cottrell JE, Boerjan W, Van BE. Gene flow between cultivated poplars and native black poplar (*Populus nigra* l.): a case study along the river Meuse on the Dutch-Belgian border. Forest Ecol Manag. 2004;197:307–10.
23. Vietto L, Vanden B, Van LK, Tautenham M, Chiarabaglio PM. Matching the needs for the European black poplar (*Populus nigra* L.) gene conservation and river restoration: case studies in Italy, Belgium and Germany. In: Proceedings of the IV ECRR International Conference on River Restoration. 2008. p. 157–66.
24. Zsuffa L. The genetics of *Populus nigra* L. Annales Forestales (Zagreb). 1974; 6:29–53.
25. Van der Schoot J, Pospíšková M, Vosman B, Smulders MJM. Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). Theor Appl Genet. 2000;101:317–22.
26. Lu P, Yan GX. Forest in Xinjiang. Urumqi: Xinjiang People's Press; 1989.
27. Yang CY, Shen KM, Mao ZM. Populus L. In: Yang CY, editor. Flora Xinjianggensis Tomus 1, vol. 2. Urumqi: Technology & Hygiene Publishing House; 1992. p. 122–58.
28. Hamzeh M, Dayanandan S. Phylogeny of *Populus* (*Salicaceae*) based on nucleotide sequences of chloroplast trnT-trnF region and nuclear rDNA. Am J Bot. 2004;91:1398–408.
29. Cervera MT, Storme V, Soto A, Ivens B, Van Montagu M, Rajora OP, et al. Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. Theor Appl Genet. 2005;111:1440–56.
30. Storme V, Broeck AV, Ivens B, Halfmaerten D, Slycken JV, Castiglione S, et al. Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development. Theor Appl Genet. 2004;108:969–81.
31. Frison E, Lefvre F, De Vries S, Turok J. *Populus nigra* network. Report of the First Meeting, 3–5 October 1994, Izmit, Turkey. Rome: IPGRI; 1995.
32. Wang C, Fang CF, Zhao SD, Chou YL, Tung SL. Populus L. In: Wang C, Fang CF, editors. Flora Republicae Popularis Sinicae Tomus, vol. 20. Beijing: Science Press; 1984. p. 1–78.
33. Feng JJ, Jiang DC, Shang HY, Dong M, Wang GN, He XY, et al. Barcoding Poplars (*Populus* L.) from Western China. Plos One. 2013;8:e71710.
34. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–59.
35. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611–20.
36. Anderson EC, Thompson EA. A model-based method for identifying species hybrids using multilocus genetic data. Genetics. 2002;160:1217–29.
37. Arnold ML. Natural hybridization and evolution. Oxford: Oxford University Press; 1997.
38. Johnston JA, Wesselingh RA, Bouck AC, Donovan LA, Arnold ML. Intimately linked or hardly speaking? The relationship between genotype and environmental gradients in a Louisiana Iris hybrid population. Mol Ecol. 2001;10:673–82.
39. Kyhos DW, Clark C, Thompson WC. The hybrid nature of *Encelia laciniata* (*Compositae: Heliantheae*) and control of population composition by post-dispersal selection. Syst Bot. 1981;6:399–411.
40. Nason JD, Ellstrand NC, Arnold ML. Patterns of hybridization and introgression in populations of oaks, manzanitas, and irises. Am J Bot. 1992;79:101–11.
41. Zha HG, Milne RI, Sun H. Asymmetric hybridisation in *Rhododendron agastum*: a hybrid taxon comprising mainly F1s in Yunnan, China. Ann Bot. 2010;105:89–100.
42. Rieseberg LH, Archer MA, Wayne RK. Transgressive segregation, adaptation, and speciation. Heredity. 1999;83:363–72.

Jiang *et al. BMC Plant Biology* (2016) 16:89

Page 12 of 12

43. Rieseberg LH, Baird SJE, Gardner KA. Hybridization, introgression, and linkage evolution. Plant Mol Biol. 2000;42:205–24.

44. Torre AR, Wang TL, Jaquish B, Aitken SN. Adaptation and exogenous selection in a *Picea glauca × Picea engelmannii* hybrid zone: implications for forest management under climate change. New Phytol. 2014;201:687–99.

45. Doyle JJ. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem bull. 1987;19:11–5.

46. Cullings KW. Design and testing of a plant specific PCR primer for ecological and evolutionary studies. Mol Ecol. 1992;1:233–40.

47. Ma T, Wang J, Zhou G, Yue Z, Hu Q, Chen Y, et al. Genomic insights into salt adaptation in a desert poplar. Nat Commun. 2013;4:657–78.

48. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2001;28:2731–9.

49. Jiang D, Wu G, Mao K, Feng J. Structure of genetic diversity in marginal populations of black poplar (populus nigra l.). Biochem Syst Ecol. 2015;61: 297–302.

50. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25:1451–2.

51. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975;7:256–76.

52. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983;105:437–60.

53. Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetic. 1985;111:147–64.

54. Depaulis F, Veuille M. Neutrality tests based on the distribution of haplotypes under an infinite-site model. Mol Biol Evol. 1998;15:1788–90.

55. Depaulis F, Mousset S, Veuille M. Haplotype tests using coalescent simulations conditional on the number of segregating sites. Mol Bio Evol. 2001;18:1136–8.

56. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123:585–95.

57. Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics. 1993;133: 693–709.

58. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes, application to human mitochondrial DNA restriction data. Genetics. 1992;131:479–91.

59. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Res. 2010;10:564–7.

60. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.

61. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Mol Ecol Res. 2009;9:1322–32.

62. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics. 2012; 28:2537–9.

63. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 2004;21:255–65.

64. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23:254–67.

65. Kilian B, Ozkan H, Deusch O, Effgen S, Brandolini A, Kohl J, et al. Independent wheat B and G genome origins in outcrossing *Aegilops* progenitor haplotypes. Mol Biol Evol. 2008;24:217–27.

66. Winkler M, Tribsch A, Schneeweiss GM, Brodbeck S, Gugerli F, Holderegger R, et al. Strong nuclear differentiation contrasts with widespread sharing of plastid DNA haplotypes across taxa in European purple saxifrages (*Saxifraga* section *Porphyrion* subsection *Oppositifoliae*). Bot J Linn Soc. 2013;173:622–36.

67. Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, et al. Inferring population history with DIYABC: a user-friendly approach to Approximate Bayesian Computations. Bioinformatics. 2008;24:2713–9.

68. Cornuet JM, Ravigné V, Estoup A. Inference on population history and model checking using DNA sequence and microsatellite data with the sofware DIYABC (v1.0). BMC Bioinformatics. 2010;11:401.

69. Petit RJ, Hampe A. Some evolutionary consequences of being a tree. Ann Rev Ecol Evol Syst. 2006;37:187–214.

70. Nielsen R, Wakeley J. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics. 2001;158:885–96.

71. Hey J, Nielsen R. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics. 2004;167:747–60.

72. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. 2006;313:1596–604.