OPEN

Genassemblage 2.0 software facilitates conservation of genetic variation of captively propagated species

Dariusz Kaczmarczyk^{1✉} & Jacek Wolnicki²

In conservation of captively propagated species, conserving genetic diversity is important. Here, we present an example of the use of Genassemblage 2.0 software in conserving the genetic variation of the lake minnow (*Eupallasella percnurus*). This fish has low genetic variation and is at risk of extinction in the western edge of its range, which includes Poland. Fish from one Polish population were captured (23 males, 25 females). Fin clips were taken, and DNA was extracted. Polymorphic microsatellites (13) were used to prepare genetic profiles, assess genetic variation in the fish and estimate genetic diversity in their progeny. Alleles were scored using an automatic capillary sequencer. The four and eight best variants of spawning pairs, and the optimal sets for group volitional breeding (four males, four females; eight males, eight females) were identified using Genassemblage 2.0. In the sets of 8 and 16 fish for group breeding, the mean heterozygosity, the number of alleles, and the share of “weak” heterozygotes (0.493, 24, 0.239 and 0.479, 23, 0.257, respectively) were better than the mean values for the progeny of all potential breeding pairs. For group volitional breeding, one set of four males and four females, and numerous sets of eight males and eight females would enable transmission of all 33 alleles identified in the potential broodstock and an expected progeny heterozygosity of 0.441 and 0.414, respectively. These expected heterozygosity values are higher than those in the broodstock. For practical purposes, the larger sets would be preferable for avoiding a future inbreeding and genetic drift.

Maintaining high genetic variation, as measured by high heterozygosity and allelic diversity, is crucial for keeping a population viable and able to adapt to environmental changes^{1–3}. Conserving genetic diversity is important in conservation of captively propagated species. Unfortunately, genetic variation within captively propagated populations may decrease^{4–6} due to (1) obtaining large groups of juveniles from a few parental individuals, (2) pairing individuals without knowledge of the genetic differences between them, (3) possible inbreeding events⁷, or (4) domestication of the broodstock⁸. A decrease in population genetic variation may reduce that population's viability and its potential to adapt to environmental stresses³. Guidelines have been developed to inform broodstock collection, mating, rearing of young, and outplanting⁹.

An example of a species that is at risk of experiencing reduced genetic variation is the lake minnow, *Eupallasella percnurus* (Pallas, 1814), a cyprinid fish. Its populations occur in the northern hemisphere from Poland to the Pacific coast¹⁰. In the western part of its range, this species is in critical danger of extinction due to the specific nature of its habitats, which are small and shallow water bodies that exist for only several decades¹¹. In Poland, the lake minnow is under strict protection, involving active conservation measures¹². One of the most common measures used so far is outplanting of juvenile fish produced in aquaculture facilities to selected water bodies to establish new populations¹³.

One way of addressing the problem of potential loss of genetic variation in such populations is by preparing genetic profiles of individuals by screening highly polymorphic fragments of DNA such as microsatellites. With these profiles, individuals that are as different as possible can be identified and assembled in to breeding pairs¹⁴. To use the information in these genetic profiles for managing genetic diversity, bioinformatic tools are needed¹⁵. To meet this need, Genassemblage 2.0 software was constructed. Its installer, a detailed user guide and examples

¹Department of Environmental Biotechnology, University of Warmia and Mazury in Olsztyn, Słoneczna 45G, 10-718 Olsztyn, Poland. ²Pond Fishery Department, Inland Fisheries Institute in Olsztyn, Oczapowskiego 10, 10-719 Olsztyn, Poland. ✉email: d.kaczmarczyk@uwm.edu.pl

population group	population	sex	sample	Z10362	Z9878	Z13419	CA3	CA4	CA12	Eupe1	Eupe2	Eupe4	Eupe5	Eupe6	Eupe7	Eupe9															
1	Guzy	M	Gu01	119	119	121	121	206	206	192	199	84	84	86	90	238	250	113	157	206	206	142	154	294	294	247	253	176	180	186	186
1	Guzy	M	Gu02	119	119	121	121	206	206	199	199	84	84	84	90	250	258	157	157	206	206	142	154	294	294	247	247	176	176	186	186
1	Guzy	F	Gu03	119	119	105	121	206	206	192	199	84	84	84	90	250	254	113	157	206	206	142	142	294	294	247	247	176	180	186	186
1	Guzy	M	Gu04	119	119	105	121	206	206	192	192	84	84	84	90	246	254	157	157	206	206	142	154	294	294	247	247	176	180	186	186
1	Guzy	M	Gu05	119	119	105	121	206	206	192	199	84	84	86	90	250	262	113	113	206	206	142	142	294	294	247	263	176	180	186	186
1	Guzy	F	Gu06	119	119	105	105	190	190	192	192	84	84	84	90	254	258	157	157	206	206	142	154	294	294	247	263	176	176	186	186
1	Guzy	F	Gu07	119	119	105	105	206	206	199	199	84	84	84	90	234	242	157	157	206	206	142	154	294	294	247	267	176	180	186	186
1	Guzy	F	Gu08	119	119	105	121	206	206	192	192	84	84	84	90	258	258	113	157	206	206	142	142	294	294	247	263	180	180	186	186
1	Guzy	F	Gu09	119	119	105	121	206	206	192	192	84	84	84	90	250	254	157	157	206	206	142	154	294	294	247	247	176	176	186	186
1	Guzy	F	Gu10	119	119	105	121	206	206	199	199	84	84	84	90	258	258	157	157	206	206	154	154	294	294	247	263	176	180	186	186
1	Guzy	M	Gu11	119	119	105	121	206	206	192	199	84	84	84	90	250	254	113	157	206	206	142	154	294	294	247	263	176	176	186	186
1	Guzy	M	Gu12	119	119	105	105	206	206	192	192	84	84	84	90	250	254	113	113	206	206	142	142	294	294	247	247	176	176	186	186
1	Guzy	M	Gu13	119	119	105	105	206	206	192	199	84	84	84	90	258	258	113	157	206	206	142	154	294	294	247	263	176	176	186	186
1	Guzy	F	Gu14	119	119	105	121	206	206	192	199	84	84	86	90	258	258	113	157	206	206	142	154	294	294	247	263	176	180	186	186
1	Guzy	F	Gu15	119	119	105	121	206	206	192	192	84	84	84	90	250	250	157	157	206	206	142	154	294	294	247	263	180	180	186	186
1	Guzy	M	Gu16	119	119	105	105	206	206	192	199	84	84	84	90	258	262	113	157	206	206	142	142	294	294	247	267	180	180	186	186
1	Guzy	M	Gu17	119	119	105	121	206	206	192	192	84	84	84	90	254	262	113	157	206	206	142	154	294	294	247	263	180	180	186	186
1	Guzy	M	Gu18	119	119	105	105	190	190	199	199	84	84	86	90	250	250	157	157	206	206	142	154	294	294	247	247	176	180	186	186
1	Guzy	M	Gu19	119	119	105	121	206	206	192	192	84	84	86	90	242	250	157	157	206	206	142	154	294	294	247	247	180	180	186	186
1	Guzy	F	Gu20	119	119	105	121	206	206	192	192	84	84	84	90	250	258	157	157	206	206	154	154	294	294	247	263	176	180	186	186
1	Guzy	F	Gu21	119	119	105	121	206	206	199	199	84	84	84	90	246	250	157	157	206	206	142	154	294	294	247	263	176	180	186	186
1	Guzy	M	Gu22	119	119	105	121	206	206	199	199	84	84	84	90	250	250	113	157	206	206	142	142	294	294	247	267	176	176	186	186
1	Guzy	M	Gu23	119	119	105	121	206	206	192	199	84	84	84	90	250	250	113	113	206	206	142	142	294	294	247	247	176	176	186	186
1	Guzy	F	Gu24	119	119	105	121	206	206	192	199	84	84	86	90	254	254	157	157	206	206	142	154	294	294	247	263	176	180	186	186
1	Guzy	F	Gu25	119	119	105	121	206	206	192	192	84	84	86	90	246	250	113	157	206	206	142	154	294	294	247	247	176	180	186	186
1	Guzy	M	Gu26	119	119	105	121	206	206	192	199	84	84	86	90	242	258	113	157	206	206	142	154	294	294	247	263	180	180	186	186
1	Guzy	M	Gu27	119	119	105	121	206	206	199	199	84	84	86	90	258	258	157	157	206	206	154	154	294	294	247	263	176	180	186	186
1	Guzy	F	Gu28	119	119	105	121	206	206	199	199	84	84	84	90	254	254	113	157	206	206	142	142	294	294	247	263	176	180	186	186
1	Guzy	M	Gu29	119	119	105	121	206	206	199	199	84	84	84	90	238	250	157	157	206	206	142	154	294	294	247	263	176	180	186	186
1	Guzy	F	Gu30	119	119	105	121	206	206	192	199	84	84	86	90	250	254	149	157	206	206	142	142	294	294	247	263	176	176	186	186
1	Guzy	M	Gu31	119	119	105	121	206	206	192	192	84	84	84	90	250	254	157	157	206	206	142	142	294	294	247	263	176	180	186	186
1	Guzy	F	Gu32	119	119	105	105	206	206	192	199	84	84	84	90	250	250	157	157	206	206	154	154	294	294	247	263	176	180	186	186
1	Guzy	F	Gu33	119	119	105	121	206	206	192	192	84	84	86	90	250	250	113	113	206	206	142	142	294	294	247	263	176	180	186	186
1	Guzy	F	Gu34	119	119	105	121	190	206	192	192	84	84	84	90	250	258	157	157	206	206	142	154	294	294	247	247	176	180	186	186
1	Guzy	M	Gu35	119	119	105	121	206	206	192	199	84	84	86	90	246	254	113	113	206	206	142	142	294	294	247	263	180	180	186	186
1	Guzy	F	Gu36	119	119	105	121	206	206	199	199	84	84	86	90	238	250	113	157	206	206	142	142	294	294	247	267	176	180	186	186
1	Guzy	F	Gu37	119	119	105	121	206	206	192	192	84	84	86	90	258	258	113	157	206	206	142	142	294	294	247	247	176	180	186	186
1	Guzy	M	Gu38	109	119	105	121	190	206	192	199	84	84	86	90	258	262	113	157	206	206	142	142	294	294	247	267	180	180	186	186
1	Guzy	F	Gu39	119	119	105	121	206	206	192	199	84	84	84	90	242	246	113	157	206	206	142	154	294	294	247	267	176	180	186	186
1	Guzy	M	Gu40	119	119	105	121	206	206	192	199	84	84	84	90	254	254	113	113	206	206	142	142	294	294	247	255	176	176	186	186
1	Guzy	M	Gu41	119	119	105	121	206	206	199	199	84	84	84	90	250	250	113	113	206	206	142	142	294	294	247	263	176	180	186	186
1	Guzy	F	Gu42	119	119	105	121	206	206	192	199	84	84	86	90	254	254	157	157	206	206	142	154	294	294	247	263	176	180	186	186
1	Guzy	F	Gu43	119	119	105	121	206	206	192	199	84	84	84	90	234	254	113	113	206	206	142	142	294	294	247	263	176	180	186	186
1	Guzy	F	Gu44	119	119	105	121	206	206	192	199	84	84	86	90	258	262	113	157	206	206	142	154	294	294	247	247	176	180	186	186
1	Guzy	M	Gu45	119	119	121	121	206	206	199	199	84	84	84	90	250	254	113	157	206	206	142	154	294	294	247	247	176	180	186	186
1	Guzy	F	Gu46	119	119	105	121	206	206	192	199	84	84	84	90	258	258	113	157	206	206	142	142	294	294	247	247	176	180	186	186
1	Guzy	M	Gu47	119	119	105	121	206	206	199	199	84	84	84	90	246	250	113	157	206	206	142	154	294	294	247	247	176	180	186	186
1	Guzy	F	Gu48	119	119	105	105	206	206	192	199	84	84	84	90	258	258	113	113	206	206	154	154	294	294	247	247	180	180	186	186

Figure 1. Genetic profiles of lake minnow individuals from the Guzy population (example of input file to Genassemblage 2.0). Column 1, population group; column 2, population name; column 3, sex; column 4, fish ID (samples); columns 5–33, names of microsatellite markers (line 1) and their alleles (lines 2–50).

SUMARISED V INDEX FOR ALL VARIANTS OF BREEDING PAIRS AND SET OF BEST PAIRS MARKED WITH BOLD																																			
FEMALE																																			
	Gu03	Gu06	Gu07	Gu08	Gu09	Gu10	Gu14	Gu15	Gu20	Gu21	Gu24	Gu25	Gu28	Gu30	Gu32	Gu33	Gu34	Gu35	Gu37	Gu39	Gu42	Gu43	Gu44	Gu46	Gu48									importance	
	Gu01	0.731	0.860	0.852	0.751	0.734	0.751	0.768	0.700	0.734	0.709	0.768	0.751	0.751	0.798	0.734	0.730	0.759	0.751	0.768	0.776	0.768	0.768	0.785	0.751	0.759	ih						0.45		
	Gu02	0.624	0.724	0.687	0.720	0.569	0.552	0.734	0.703	0.637	0.573	0.700	0.717	0.654	0.747	0.636	0.797	0.623	0.734	0.683	0.709	0.700	0.755	0.684	0.603	0.690	iwh						0.1		
	Gu04	0.764	0.612	0.775	0.691	0.657	0.724	0.725	0.589	0.657	0.696	0.674	0.741	0.724	0.747	0.691	0.797	0.680	0.826	0.758	0.764	0.674	0.758	0.826	0.775	0.797	iar						0.45		
	Gu05	0.680	0.859	0.852	0.700	0.767	0.817	0.767	0.733	0.801	0.741	0.801	0.751	0.700	0.764	0.767	0.628	0.792	0.734	0.717	0.776	0.801	0.667	0.751	0.700	0.708							$\Sigma=1.0$		
	Gu11	0.640	0.707	0.772	0.737	0.602	0.704	0.751	0.686	0.687	0.674	0.700	0.734	0.653	0.713	0.653	0.730	0.725	0.768	0.751	0.742	0.700	0.687	0.768	0.704	0.725									
	Gu12	0.539	0.639	0.770	0.636	0.534	0.804	0.784	0.702	0.703	0.741	0.733	0.633	0.636	0.712	0.669	0.578	0.657	0.767	0.599	0.708	0.733	0.586	0.717	0.602	0.606									
	Gu13	0.691	0.623	0.704	0.636	0.653	0.602	0.649	0.720	0.653	0.691	0.733	0.751	0.686	0.764	0.636	0.763	0.691	0.785	0.649	0.725	0.733	0.704	0.700	0.602	0.573									
	Gu16	0.657	0.775	0.670	0.586	0.754	0.720	0.751	0.704	0.754	0.742	0.785	0.768	0.687	0.865	0.720	0.764	0.725	0.751	0.649	0.708	0.785	0.704	0.700	0.602	0.573									
	Gu17	0.674	0.741	0.805	0.602	0.670	0.737	0.751	0.602	0.670	0.741	0.700	0.717	0.686	0.831	0.704	0.730	0.708	0.835	0.700	0.742	0.700	0.687	0.734	0.704	0.641									
	Gu18	0.707	0.696	0.716	0.851	0.716	0.716	0.880	0.749	0.767	0.679	0.796	0.796	0.767	0.847	0.632	0.830	0.679	0.814	0.830	0.831	0.796	0.868	0.814	0.767	0.746									
	Gu19	0.657	0.780	0.750	0.649	0.615	0.733	0.814	0.581	0.633	0.713	0.730	0.662	0.784	0.830	0.649	0.712	0.629	0.848	0.679	0.730	0.730	0.801	0.747	0.700	0.679									
MALE	Gu22	0.595	0.780	0.640	0.723	0.623	0.690	0.758	0.707	0.741	0.612	0.741	0.707	0.606	0.696	0.657	0.679	0.713	0.607	0.690	0.680	0.741	0.674	0.725	0.623	0.729									
	Gu23	0.539	0.757	0.770	0.669	0.585	0.753	0.767	0.702	0.720	0.673	0.767	0.649	0.636	0.712	0.669	0.578	0.673	0.700	0.632	0.691	0.767	0.602	0.700	0.585	0.639									
	Gu26	0.815	0.859	0.734	0.683	0.852	0.734	0.717	0.700	0.768	0.776	0.801	0.835	0.785	0.865	0.785	0.814	0.809	0.785	0.767	0.742	0.801	0.802	0.801	0.751	0.675									
	Gu27	0.797	0.780	0.730	0.746	0.763	0.527	0.679	0.746	0.662	0.662	0.712	0.814	0.746	0.801	0.662	0.850	0.746	0.797	0.746	0.798	0.712	0.847	0.730	0.712	0.646									
	Gu29	0.674	0.775	0.670	0.737	0.670	0.602	0.751	0.636	0.670	0.573	0.700	0.767	0.653	0.764	0.602	0.796	0.707	0.683	0.784	0.742	0.700	0.754	0.768	0.704	0.725									
	Gu31	0.589	0.657	0.754	0.602	0.552	0.720	0.751	0.552	0.619	0.657	0.649	0.683	0.636	0.679	0.636	0.662	0.623	0.751	0.649	0.742	0.649	0.670	0.768	0.653	0.758									
	Gu35	0.680	0.859	0.852	0.649	0.801	0.817	0.767	0.733	0.835	0.741	0.750	0.751	0.649	0.797	0.817	0.679	0.826	0.768	0.717	0.742	0.750	0.633	0.785	0.700	0.657									
	Gu38	0.815	0.877	0.869	0.734	0.903	0.869	0.869	0.852	0.903	0.877	0.903	0.886	0.835	0.966	0.919	0.865	0.809	0.869	0.767	0.843	0.903	0.853	0.818	0.751	0.759									
	Gu40	0.589	0.741	0.788	0.703	0.636	0.787	0.784	0.804	0.804	0.758	0.717	0.751	0.586	0.730	0.787	0.696	0.775	0.768	0.683	0.725	0.717	0.586	0.751	0.636	0.690									
	Gu41	0.595	0.830	0.724	0.656	0.707	0.689	0.707	0.656	0.740	0.612	0.741	0.707	0.572	0.713	0.656	0.594	0.746	0.607	0.690	0.714	0.741	0.589	0.725	0.623	0.646									
Gu45	0.595	0.814	0.724	0.724	0.623	0.640	0.758	0.707	0.707	0.612	0.674	0.691	0.589	0.780	0.673	0.746	0.696	0.708	0.707	0.697	0.674	0.674	0.691	0.640	0.663										
Gu47	0.623	0.825	0.704	0.737	0.670	0.653	0.784	0.720	0.720	0.589	0.751	0.683	0.653	0.831	0.653	0.763	0.707	0.734	0.733	0.674	0.751	0.721	0.717	0.653	0.641										
		H	wh	ar	v																														
	all pairs average	0.377	0.364			21	0.720																												
	4 pairs	average	0.493	0.263			24	0.890																											
	8 pairs	average	0.479	0.257			23	0.868																											
	4 pairs	chi2	0.1884																																
	8 pairs	chi2	0.1705																																

Figure 2. The sets of four and eight best breeding pairs identified by using the ν index. ID of males are given in column 2, and those of females in line 3. Values of the ν index are given in lines 4–26 and columns 3–28. The names males and females included in for sets of four and eight pairs are marked in light grey. Their ν values are marked in bold. The males and females included in set of eight best pairs only are marked in dark grey, their ν values are marked in bold and underlined. The mean values of H , wh , ar , and ν index for selected pairs is compared to the means for all pairs in lines 29–31. The importance of each component of the ν index is given in column 31. The χ^2 values are given in line 33–34.

heterozygotes (i_{wh}) 0.1, and a relative index of allelic number (i_{ar}) 0.45; the number of breeding pairs was set at four and then at eight.

To find the optimal set for volitional group-spawning, the module *Find best set of individuals for group spawning* was used. One set consisted of four males and four females, and the other, of eight males and eight females. The number of alleles was selected as the primary indicator. The calculation performed by this module is based on the assumption that each spawner contributes equally to the progeny.

Ethical statement. These studies were approved by the Local Committee for the Experiments on Animals in Olsztyn, Poland (permission No. 15/2015) issued on 25.03.2015. All procedures were performed in accordance to relevant guidelines.

Results

In the group of broodstock candidates, 33 alleles were identified across all microsatellite loci. The H_o was 0.386, H_e was 0.375, and the share of weak heterozygotes was 0.208.

With this group of male and female candidates, it was possible to assemble breeding pairs in 575 different combinations. The progeny of each potential pair would have inherited from 17 to 26 (mean of 21) alleles. This is lower than the 33 alleles that were identified in the parental group. Across the investigated markers, heterozygosity ranged from 0.192 to 0.519 (mean 0.378), but only in 211 pairs (36.7%) was the observed heterozygosity higher than that in the parental group as a whole. The average share of weak heterozygotes in groups of progeny was 0.365, ranging from 0.167 to 0.500. The value of this indicator was less than or equal to that in the parental group in only 65 progeny groups (11.3%).

Based on the ν index, a set of four pairs were selected as the optimal pairs for breeding: male Gu01 and female Gu06, male Gu05 and female Gu07, male Gu18 and female Gu14, and male Gu38 and female Gu30 (Fig. 2). The average heterozygosity of this set was 0.493, the number of alleles was 24, the share of weak heterozygotes was 0.263, and the ν index of this set was 0.890. The ν index also was used to select a set of the eight best pairs. This set included the four same pairs as listed above and four others that were added to set of four males and four females. They were: male Gu19 and female Gu36, male Gu26 and female Gu09, male Gu27 and female Gu33, and male Gu35 and female Gu20. The average heterozygosity of this set was 0.479, the number of alleles was 23, and the share of weak heterozygotes was 0.257. The ν index of this set was 0.868 (Fig. 2). All these values of genetic diversity indicators were higher than the averages for all potential breeding pairs (0.377, 21, 0.364, and 0.720, respectively).

For group volitional breeding, the 23 males could have been assembled into 8855 different combinations of four individuals, and the 25 females, into 12,650 such combinations. Together, over 120 million different

	Z10362	Z9878	Z13419	CA3	CA4	CA12	Eupe1	Eupe2	Eupe4	Eupe5	Eupe6	Eupe7	Eupe9	acros loci
nr of alleles	2	2	2	2	3	8	3	1	2	1	4	2	1	33
heterozygosity	0.125	0.563	0.313	0.500	0.950	0.860	0.719	0.000	0.406	0.000	0.734	0.563	0.000	0.441
sex	sample													
F	Gu06													
F	Gu07													
F	Gu30													
F	Gu39													
M	Gu01													
M	Gu35													
M	Gu38													
M	Gu40													

Figure 3. The set of four males and four females for volitional breeding. Line 1—the names of microsatellite markers, lines 2–3—the number of alleles inherited by progeny of this set and their heterozygosity at each locus. Column 30—mean heterozygosity and number of alleles across loci are in bold. Lines 5–13 fish IDs (samples) included in this set (column 2) and their sex (column 1).

	Z10362	Z9878	Z13419	CA3	CA4	CA12	Eupe1	Eupe2	Eupe4	Eupe5	Eupe6	Eupe7	Eupe9	acros loci
nr of alleles	2	2	2	2	3	8	3	1	2	1	4	2	1	33
heterozygosity	0.063	0.500	0.313	0.508	0.953	0.883	0.559	0.000	0.438	0.000	0.656	0.516	0.000	0.414
sex	sample													
F	Gu06													
F	Gu07													
F	Gu08													
F	Gu10													
F	Gu30													
F	Gu36													
F	Gu39													
F	Gu44													
M	Gu01													
M	Gu04													
M	Gu05													
M	Gu17													
M	Gu18													
M	Gu35													
M	Gu38													
M	Gu40													

Figure 4. The set of eight males and eight females for volitional breeding. Line 2—the names of microsatellite markers, lines 3–4—the number of alleles inherited by progeny of this set and their heterozygosity at each locus. Column 30—mean heterozygosity and number of alleles across loci are in bold. Lines 6–22 fish IDs (samples) included in this set (column 2) and their sex (column 1).

combinations of males and females could have been created. The best set included the males Gu0, Gu35, Gu38 and Gu40, and the females Gu06, Gu07, Gu30, and Gu39. This group would have enabled the transmission of all 33 alleles and in the presented stock, assuming equal contribution, a heterozygosity value of 0.441 in their progeny (Fig. 3). Thus, this set would have transmitted all of the allelic diversity identified in the group of 48 fish sampled from the Guzy population, and the heterozygosity in their progeny would have been higher than that in the parental group. In addition, 2083 sets of individuals were identified that would have transferred 32 or 33 alleles and whose progeny would have had heterozygosity values ranging from 0.400 to 0.440. Those sets were potential alternatives to the set that we selected. When the number of fish in the set was increased to eight males and eight females, the number of potential sets of females increased to 490,314, and that of males, to 1,081,575. Together, this resulted in 5.30311×10^{11} combinations for this set. Among them, a large number of combinations enabled inheritance of all 33 alleles and an average heterozygosity of progeny of 0.414. An example of such a set is given in Fig. 4.

Discussion

Identification of a group of pairs or sets of group-spawning individuals that are optimal for conservation of genetic variation can be time-consuming and difficult. Manually finding the best of them is problematic because of the complexity of the calculations and the huge number of potential pair variants or set compositions. For this reason, bioinformatic tools can be very helpful. Genassemblage 2.0 is a versatile tool for artificial breeding of conserved species such as the lake minnow. It can be used to identify the best breeding set for either individual or group breeding. This is merely an example of how the software works (although the number of markers should be high enough to track genetic differences both within and between lake minnow populations^{16,22}). If any molecular marker is polymorphic and inherited according to Mendelian laws, and numbers can be assigned its alleles, then it can be used. Genassemblage software does not have a limit at the number of molecular

markers that can be used. In this study, we used the polymorphism of microsatellite fragments as a marker of genetic differences between individuals. Markers other than microsatellite fragments such as single nucleotide polymorphism can be used, if numerical values are assigned to their alleles.

Using the example of lake minnow individuals from the Guzy population, we have shown that it is possible to identify genetic differences in the broodstock and find the optimal set of spawning pairs. In Poland, the lake minnow requires active protection measures, which can be accomplished without a decrease in the genetic variation of this species by initiating (or supplementing) existing populations with juveniles originating from aquaculture facilities^{13,15,23,24}. To prevent decreases in genetic variation and obtain the most genetically diverse set of juveniles, breeding pairs need to be assembled appropriately.

In the group of potential spawners from the Guzy population, we found that the values of genetic variation indicators, such as heterozygosity, allelic diversity and share of weak heterozygotes, indicate that genetic variation in the group of potential parents is low. This finding is similar to reports on other Polish populations of this species¹⁶.

Our results indicate that random selection of fish for breeding is likely to result in lower genetic variation in the offspring than in the parents. It is known that using a small number of individuals for breeding without information about their genetic similarity and relationships (random selection) can accelerate a decrease in allelic diversity^{25–27}. Consequently decrease of population fitness in the next generation or generations is more likely²⁸. In the presented example, we found that using a set of breeding pairs chosen by Genassemblage 2.0 will provide a higher level of genetic variation in the progeny than the mean value obtained with the progeny of all potential breeding pairs. Identification of the best set of breeding pairs using the ν index can increase heterozygosity and reduce the number of weak heterozygotes in progeny. Moreover, each individual included in the set of breeding pairs proposed by Genassemblage 2.0 appears in only one breeding pair, which reduces the risk of inbreeding in future generations.

It should be noted that, in the example presented above, four breeding pairs were chosen for illustrative purposes only. In a real-world application, it would be better to choose more individuals. Using only a few individuals for breeding is associated with a higher risk of a decrease in N_e ²⁷, inbreeding in future generations and increased genetic drift in the population. It is difficult to indicate exactly how many pairs should be used for breeding, but larger numbers help to conserve within-population genetic variation⁹. Consequently, the example that was presented with eight breeding pairs would be more appropriate for conservation of genetic variation than the set of four pairs, even if the four pairs had slightly better genetic variation indicators. Generally speaking, if enough males or females are available and the same individual is not used more than once, it is recommended to mate more pairs than four and use Genassemblage 2.0 software to help find the best ones. Genassemblage 2.0 software can be helpful for volitional breeding. If we can assume that the contribution of each spawner to the progeny is equal or very similar, we can find a set of individuals that will transmit all the alleles identified in a group of potential parents to their progeny. In this situation, it is even possible to obtain progeny with higher heterozygosity than in the parents' generation. Moreover, the software can indicate alternative sets for volitional breeding that provide a level of genetic diversity similar to that in the best set. This can be useful when it is not possible to use the best set. It should be taken into account that, in volitional mass spawning of some fish species, the parental contribution to the progeny can be highly skewed^{29–35}. However, this has not been demonstrated with the lake minnow, and our results suggest that using larger sets of males and females in volitional breeding could be more suitable than using separate pairs. Again, it should be noted that, although the average predicted heterozygosity would be slightly lower with eight males and eight females than with four of each, it would be better to use the larger set for the reasons given in the preceding paragraph.

The previous version of the software (Genassemblage 1.0) accurately predicted genetic variation in the progeny of breeding pairs³⁶. In comparison, Genassemblage 2.0 has expanded functionality and can be used for more diverse tasks, such as group breeding. Moreover, the calculation of the ν index, which was implemented in Genassemblage 2.0, enables comparison and evaluation of each breeding pair and selection of the best one. Genassemblage 2.0 has a modular interface, expanded functionality and a more user-friendly design, which makes it even more useful and easy-to-use in conservation of genetic variation. Importantly, the use of Genassemblage 2.0 software is not limited to fish species. It can be used to manage the genetic variation of most species for both conservation and commercial purposes.

Received: 1 April 2020; Accepted: 5 October 2020

Published online: 21 October 2020

References

1. Lacy, R. C. Importance of genetic variation to the viability of mammalian populations. *J. Mamm.* **78**, 320–335 (1997).
2. Hansen, M. M., Villanueva, B., Nielsen, E. E. & Bekkevold, D. Investigating the genetics of populations. In *The Atlantic Salmon: Genetics, Conservation and Management* (eds Verspoor, E. et al.) 86–114 (Blackwell Publishing, Hoboken, 2008).
3. Hallerman, E. Inbreeding. In *Population Genetics: Principles and Applications for Fisheries Scientists* (ed. Hallerman, E. M.) (American Fisheries Society, Bethesda, 2003).
4. Koljonen, M. L., Jansson, H., Paaver, T., Vasin, O. & Koskineniemi, J. Phylogeographic lineages and differentiation pattern of Atlantic salmon in the Baltic Sea with management implications. *Can. J. Fish. Aquat. Sci.* **56**, 1766–1780 (1999).
5. Koljonen, M. L., Tähtinen, J., Säisä, M. & Koskineniemi, J. Maintenance of genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding programs and the geographic distribution of microsatellite variation. *Aquaculture* **212**, 69–92 (2002).
6. Verspoor, E. et al. Population structure in the Atlantic salmon: Insights from 40 years of research into genetic protein variation. *J. Fish. Biol.* **67**(Suppl A), 3–55 (2005).
7. Bryant, E. H., McCommas, S. A. & Combs, L. M. The effect of an experimental bottleneck upon quantitative genetic variation in the housefly. *Genetics* **114**, 1191–1211 (1986).

8. Kim, J. E., Withler, R. E., Ritland, C. & Cheng, K. M. Genetic variation within and between domesticated Chinook salmon, *Oncorhynchus tshawytscha*, strains and their progenitor populations. *Dev. Environ. Biol. Fish.* **23**, 371–378 (2004).
9. Miller, L. M. & Kapuscinski, A. R. Genetic guidelines for hatchery supplementation programs. In *Population Genetics: Principles and Applications for Fisheries Scientists* (ed. Hallerman, E. M.) 329–355 (American Fisheries Society, Bethesda, 2003).
10. Kuszniierz, J., Paśko, Ł. & Tagayev, D. On the variation and distribution of the lake minnow, *Eupallasella percunurus* (Pall.). *Arch. Pol. Fish.* **19**, 161–166 (2011).
11. Sowińska-Swierkosz, B. & Kolejko, M. Extinction risk to lake minnow (*Eupallasella percunurus*) due to habitat loss: Eastern Poland case study. *Environ. Monit. Assess.* **191**, 571 (2019).
12. Kaczmarczyk, D., Wolnicki, J., Kamiński, R., Sikorska, J. & Radtke, G. Genetic variability of the endangered fish lake minnow, *Eupallasella percunurus* (Pall.) in populations newly established by translocations in Poland. in *XVI European Congress of Ichthyology Lausanne, Swiss, 2–6 September 2019. Frontier of Marine Science, Conference Abstract: XVI European Congress of Ichthyology*, 92–93.
13. Wolnicki, J., Kamiński, R., Sikorska, J. & Kaczmarczyk, D. Occurrence and active protection of the endangered cyprinid fish species, lake minnow *Eupallasella percunurus* (Pall.), in Poland. in *XVI European Congress of Ichthyology Lausanne, Swiss, 2–6 September 2019. Frontier of Marine Science, Conference Abstract: XVI European Congress of Ichthyology*, 206–208.
14. Attard, C. *et al.* A novel holistic framework for genetic-based captive-breeding and reintroduction programs. *Conserv. Biol.* **30**, 1060–1069 (2016).
15. Kaczmarczyk, D. Techniques based on the polymorphism of microsatellite DNA as tools for conservation of endangered populations. *Appl. Ecol. Environ. Res.* **17**, 1599–1615 (2019).
16. Kaczmarczyk, D. & Wolnicki, J. Genetic diversity of the endangered cyprinid fish lake minnow *Eupallasella percunurus* in Poland and its implications for conservation. *PLoS ONE* **12**, 1–16 (2016).
17. Dimsoski, P., Toth, G. P. & Bagley, M. J. Microsatellite characterization in central stoneroller *Camptostoma anomalum* (Pisces: Cyprinidae). *Mol. Ecol.* **9**, 2187–2189 (2000).
18. Holmen, J., Vøllestad, L. A., Jakobsen, K. S. & Primmer, C. R. Cross-species amplification of zebrafish and central stoneroller microsatellite loci in six other cyprinids. *J. Fish. Biol.* **66**, 851–859 (2005).
19. Aksoy, S. *et al.* Permanent genetic resources. *Mol. Ecol. Res.* **13**, 341–343 (2013).
20. Dieringer, D. & Schlötterer, C. Microsatellite analyzer (MSA): A platform independent analysis tool for large microsatellite data sets. *Mol. Ecol. Notes* **3**, 167–169 (2003).
21. Kaczmarczyk, D. & Fopp-Bayat, D. Assemblage of spawning pairs based on their individual genetic profiles—As tool for maintaining genetic variation within sturgeon populations. *Aquacult. Res.* **44**, 677–682 (2013).
22. Ruzzante, D. E. A comparison of several measures of genetic distance and population structure with microsatellite data: Bias and sampling variance. *Can. J. Fish. Aquat. Sci.* **55**, 1–14 (1998).
23. Kamiński, R., Kuszniierz, J., Myszkowski, L. & Wolnicki, J. The first attempt to artificially reproduce the endangered cyprinid lake minnow *Eupallasella percunurus* (Pallas). *Aquacult. Int.* **12**, 3–10 (2004).
24. Wolnicki, J., Kamiński, R. & Sikorska, J. Occurrence, threats and active protection of the lake minnow, *Eupallasella percunurus* (Pall.), in Mazowieckie Voivodeship in Poland. *Arch. Pol. Fish.* **19**, 209–216. <https://doi.org/10.2478/v10086-011-0026-3> (2011).
25. Liu, D. *et al.* Low genetic diversity in broodstocks of endangered Chinese sucker, *Myxocyprinus asiaticus*: Implications for artificial propagation and conservation. *ZooKeys* **792**, 117–132 (2018).
26. Hariyono, D. N. H. *et al.* Genetic diversity and phylogenetic relationship analyzed by microsatellite markers in eight Indonesian local duck populations. *Asian-Australas J. Anim. Sci.* **32**, 31–37 (2019).
27. Ryman, N. & Laikre, L. Effects of supportive breeding on the genetically effective population size. *Conserv. Biol.* **5**, 325–329 (1991).
28. Hansen, M. M., Nielsen, E. E., Ruzzante, D. E., Bouza, C. & Mensberg, K. Genetic monitoring of supportive breeding in brown trout *Salmo trutta* L., using microsatellite DNA markers. *Can. J. Fish. Aquat. Sci.* **57**, 2130–2139 (2000).
29. Bartley, D., Bagley, M., Gall, G. & Bentley, B. Use of linkage disequilibrium data to estimate effective size of hatchery and natural fish populations. *Conserv. Biol.* **6**, 365–375 (1992).
30. Hara, M. & Sekino, M. Efficient detection of parentage in a cultured Japanese flounder *Paralichthys olivaceus* using microsatellite DNA marker. *Aquaculture* **217**, 107–114 (2003).
31. Sekino, M. *et al.* Microsatellite-based pedigree tracing in a Japanese flounder *Paralichthys olivaceus* hatchery strain: Implications for hatchery management related to stock enhancement program. *Aquaculture* **221**, 255–263 (2003).
32. Brown, R. C., Woolliams, J. A. & McAndrew, B. J. Factors influencing effective population size in commercial populations of gilthead seabream, *Sparus aurata*. *Aquaculture* **247**, 219–225 (2005).
33. Frost, L. A., Evans, B. S. & Jerry, D. R. Loss of genetic diversity due to hatchery culture practices in barramundi (*Lates calcarifer*). *Aquaculture* **261**, 1056–1064 (2006).
34. Loughnan, S. R. *et al.* Broodstock contribution after mass spawning and size grading in barramundi (*Lates calcarifer*, Bloch). *Aquaculture* **404–405**, 139–149 (2013).
35. Herlin, M. *et al.* Analysis of the parental contribution to a group of fry from a single day of spawning from a commercial Atlantic cod (*Gadus morhua*) breeding tank. *Aquaculture* **274**, 218–224 (2008).
36. Kaczmarczyk, D., Dobosz, S. & Kaczor, A. Prediction of genetic variation in stocks of offspring and its correlation with viability and growth rate. *J. World Aquacult. Soc.* **1**, 1–8 (2017).

Acknowledgements

These studies were financed by the National Centre of Science (Poland) within project No. 2014/15/B/NZ9/05240 for the years 2015–2019. Project financially supported by Minister of Science and Higher Education in the range of the program entitled “Regional Initiative of Excellence” for the years 2019–2022, Project No. 010/RID/2018/19, amount of funding 12.000.000 PLN.

Author contributions

Study design: D.K. Obtaining financial support for study: J.W. and D.K. Material: J.W. Genetic and Bioinformatic Analysis: D.K. Preparation of manuscript: D.K. and J.W.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020