NPAFC
Doc. 907
Rev. _____

SNPs provide an Easily-Standardized Baseline for NPAFC Studies of Chum Salmon

by

L. W. Seeb[1], W. D. Templin[1], C. T. Smith[1], C. Elfstrom[1], S. Urawa[2], R. L. Wilmot[3], S. Abe[4] and J. E. Seeb[1]

[1]*Gene Conservation Laboratory, Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518*
[2]*Genetics Section, National Salmon Resources Center, 2-2 Nakanoshima, Toyohira-ku, Sapporo 062-0922, Japan*
[3]*National Marine Fisheries Service, Auke Bay Laboratory, 11305 Glacier Highway, Juneau, AK, USA 99801*
[4]*Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan*

Submitted to the
NORTH PACIFIC ANADROMOUS FISH COMMISSION
BY THE
UNITED STATES OF AMERICA

October, 2005

THIS PAPER MAY BE CITED IN THE FOLLOWING MANNER:
Seeb, L. W, W. D. Templin, C. T. Smith, C. Elfstrom, S. Urawa, R. L. Wilmot, S. Abe, and, J. E. Seeb. 2005. SNPs provide an Easily-Standardized Baseline for NPAFC Studies of Chum Salmon (NPAFC Doc. 907) 12 p. Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518.

Abstract

Genetic stock identification using a comprehensive allozyme baseline has been extremely valuable for high-seas and coastal migration studies from throughout the Pacific Rim. However, newer DNA techniques provide significant advantages over allozymes in sampling, sample handling, and the potential for improved resolution. Of the various DNA markers, single nucleotide polymorphisms (SNPs) assayed through high-throughput technologies are particularly appropriate for NPAFC applications. Unlike marker types based on fragment size, SNPs are based on the actual DNA sequence and require no inter-laboratory standardization. SNP data can be easily transferred between laboratories and instrument platforms. Here we present the first SNP baseline appropriate for NPFAC applications. The baseline includes 36 SNPs with both nuclear and mtDNA SNPs. The mtDNA SNPs are based on sequences previously shown to be useful for NPAFC work. A total of 73 populations ranging from Japan to the Pacific Northwest were assayed. The results demonstrate that SNPs can accurately discriminate among the regional lineages of chum salmon important for NPAFC work. SNPs are a simple and cost effective method that could be readily incorporated into NPAFC laboratories for BASIS and similar high-seas applications.

Background

Genetic stock identification studies have become a central part of North Pacific Anadromous Fish Commission (NPAFC) research activities. The NPAFC *ad hoc* Working Group on Stock Identification Studies has coordinated these studies with the goals to develop, standardize, and disseminate genetic databases among the parties. These databases must provide an appropriate level of accuracy and precision for stock identification studies while at the same time be easily shared and repeatable among the NPAFC Parties.

Genetic databases on chum salmon, in particular, have been coordinated through NPAFC. During the 1990's, large data sets of allele frequencies from nuclear allozyme loci originating from Washington, British Columbia, Alaska, Canada, Russia, and Japan were completed. These data were collected by numerous agencies from around the Pacific Rim. The database was coordinated and disseminated in recent NPAFC documents (Kondzela et al. 2002) as well as in Seeb et al. (2004a). The allozyme databases met the two goals of NPAFC genetic databases: sufficient accuracy as well as repeatability across laboratories and nations. The database has been widely used by NPAFC scientists and continues to be used to track the migration of juvenile chum salmon in the Gulf of Alaska and Bering Sea (Urawa et al. 2000; Kondzela and Wilmot 2002) and in the Bering/Aleutian Salmon International Survey (BASIS) (e.g. Urawa et al. 2004).

Development of a DNA Database

Concurrently with the application of the allozyme database, genetic markers based on DNA variation including mitochondrial DNA (mtDNA) PCR-RFLP, mtDNA sequence analyses, and microsatellite markers were developed (e.g. Park et al. 1993; Scribner et al. 1998; Seeb and Crane 1999a; Abe et al. 2002; Beacham et al. 2004). However, developing high-throughput techniques has limited the applicability of some DNA marker types (e.g. mtDNA PCR-RFLP and mtDNA sequence analyses) while difficulties in transferability and repeatability across laboratories and instrument platforms have inhibited others (e.g. microsatellite analyses). This document reports the development of an NPAFC database based on single nucleotide polymorphisms (SNPs) assayed through high-throughput technologies which are particularly conducive for NPAFC applications.

Single nucleotide polymorphisms (SNPs) have been genotyped in salmon using a broad range of techniques (e.g. Park et al. 1993; Cronin et al. 1993). Although many of these SNPs were identified over a decade ago, the chemistries available for applying these markers were relatively slow and cumbersome. Recent developments have produced rapid SNP genotyping technologies which are now faster than those for any other marker class (Moriya et al. 2004; Smith et al. 2004). SNP data can be standardized to external DNA sequences and are thus automatically standardized across chemistries, instrument platforms, and laboratories.

SNP Baseline Development

Previous reports described the initial efforts to develop SNP loci for chum salmon. Smith et al. (2005) report 13 SNPs originating from five nuclear and two mitochondrial DNA (mtDNA) sequences and included the polymorphic mitochondrial DNA (mtDNA) SNPs previously described by Sato et al. (2004). Those 13 SNPs were evaluated for their applicability for high-seas studies (Seeb et al. 2004b) using eight populations ranging from Japan to Washington State.

Here we describe an expanded database of over 6,200 individuals from 73 populations from throughout the Pacific Rim ranging from Japan to Washington State (Table 1). These populations included representatives of nearly all the known lineages of chum salmon (Seeb et al. 2004a). All individuals were assayed for 36 SNPs of which 30 were nuclear and six were mtDNA (Table 2). SNPs from mtDNA were combined and treated as a single locus resulting in a 31 locus analysis.

Results

An unrooted UPGMA tree based on genetic chord distances (Cavalli-Sforza and Edwards 1967) was constructed (Figure 1). Relationships depicted in the tree closely follow those previously revealed by allozyme data (Seeb and Crane 1999b; Seeb et al. 2004a). Japanese populations were the most divergent with Gulf of Alaska and Pacific Northwest populations clustering closely. Similar to allozymes, Susitna River populations cluster with Northwest Alaska populations rather than Gulf of Alaska populations. Russian and Chinese populations cluster in a position intermediate between Japanese and Alaskan populations.

We conducted simulation analyses using SPAM version 3.7 (Debevec et al. 2000) to test the power of the SNP markers to discriminate among the representative populations (Figure 2). Conditional maximum likelihood estimates of population contributions were made for 1000 artificial mixtures with N=400 individuals. The expected mean contribution was equal to 100%. Mean contribution estimates were greater than 90% to continent of origin, to the four major regions (Japan, Russia, Alaska, and Washington), and the majority of the 15 more localized reporting regions. Only South Alaska Peninsula (87%) and Southern Southeast Alaska (85%) fell below 90%. In both cases, the majority of the misallocation was to an adjacent region.

In addition to the 100% simulations, two additional simulations were conducted in which the ability of the baseline to identify single populations as well as correct regional memberships was examined (Table 3). Mixtures 1 includes broad representation from both Asia and North America; populations in mixture 2 are limited to those originating from western Alaska and the Canadian portion of the Yukon River. Estimated stock compositions were typically within 4-5 % of the expected value for the specific population and within 2-3% of expected to the correct reporting region.

Conclusions

The results presented here demonstrate that SNPs can accurately discriminate among the regional lineages of chum salmon important to NPAFC studies. Fine-scale differentiation within regions was also demonstrated. Since the markers directly reflect the underlying DNA sequences, they are readily repeatable and combined with data collected across hardware, chemistry platforms, and laboratories. SNP data can be generated rapidly and are well suited to mixture and migratory studies of chum salmon for a variety of Pacific Rim and NPAFC applications. This is particularly important in the context of multi-jurisdictional fisheries management where both accuracy and repeatability are required.

Literature Cited

Abe, S., Sato, S., Kojima, H., Ando, J., Anod, H., Wilmot, R.L., Seeb, L.W., Efremov, V., Leclair, L., Buchholtz, W., Jin, D-H., Urawa, S., Kaeriyama, M., Urano, A. 2002. Genetic differentiation among Pacific Rim populations of chum salmon inferred from mitochondrial DNA sequence variation. (NPAFC Doc 616) 27 pages. Laboratory of Animal Cytogentics, Center for Advanced Science and Technology, Hokkaido University, North 10, West 8, Kita-ku, Saporo 060-0810, Japan.

Beacham, T. D., K. D. Le, and J. R. Candy. 2004. Population structure and stock identification of chum salmon (*Oncorhynchus keta*) based upon microsatellite analysis. N. Pac. Anadr. Fish Comm. Tech. Rep 5: 31-33.

Cavalli-Sforza, L. L. and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Evolution 21: 550-570.

Cronin, M. A., W. J. Spearman, R. L. Wilmot, J. C. Patton and J. W. Bickham. 1993. Mitochondrial DNA variation in chinook (*Oncorhynchus tshawytscha*) and chum salmon *(O. keta*) detected by restriction enzyme analysis of polymerase chain reaction (PCR) products. Canadian Journal of Fisheries and Aquatic Sciences 50: 708-715.

Debevec, E. M., R. B. Gates, M. Masuda, J. Pella, J. Reynolds, and L. W. Seeb. 2000. SPAM (Version 3.2): Statistics program for analyzing mixtures. Journal of Heredity 91: 509-511.

Kondzela, C.M., P. A. Crane, S. Urawa, N. V. Varnavskaya, V. Efremov, X. Luan, W. D. Templin, K. Hayashizaki, R. L. Wilmot, and L. W. Seeb. 2002. Development of a comprehensive allozyme baseline for Pacific Rim chum salmon. (NPAFC Doc. 629) 23 pages. Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518.

Kondzela, C. and R. Wilmot. 2002. Origin of juvenile chum salmon from Gulf of Alaska coastal waters, 2001. Alaska Fishery Science Center Quarterly Report. July-August-September, 9 pp.

Moriya, S., A. Urano, S. Urawa, O. Suzuki, and S. Abe. 2004. Development of DNA microarray for rapid detection of mitochondrial DNA haplotypes of chum salmon. N. Pac. Anadr. Fish Comm. Tech. Rep. 5 28-30.

Park, L. K., M. A. Brainard, D. A. Dightman, G. A. Winans. 1993. Low levels of intraspecific variation in the mitochondrial DNA of chum salmon (*Oncorhynchus keta*). Molecular Marine Biology and Biotechnology 2: 362-370.

Sato, S., H. Kojima, J. Ando, H. Ando, R. L.Wilmot, L. W. Seeb, V. Efremov, L. Leclair, W. Buchholz, D. H. Jin, S. Urawa, M. Kaeriyama, A. Urano, and S. Abe. 2004. Genetic population structure of chum salmon in the Pacific Rim inferred from mitochondrial DNA sequence variation. Environmental Biology of Fishes 69: 37-50.

Scribner, K. T., P. A. Crane, W. J. Spearman, and L. W. Seeb. 1998. DNA and allozyme markers provide concordant estimates of population differentiation: Analyses of U.S. and Canadian populations of Yukon River fall-run chum salmon (*Oncorhynchus keta*). Canadian Journal of Fisheries and Aquatic Sciences 55: 1748-1758

Seeb, L. W. and P. A. Crane. 1999a. Allozymes and mitochondrial DNA discriminate Asian and North American populations of chum salmon in mixed-stock fisheries along the south coast of the Alaska Peninsula. Transactions American Fisheries Society 128: 88-103.

Seeb, L. W. and P. A. Crane. 1999b. High genetic heterogeneity in chum salmon in Western Alaska, the contact zone between northern and southern lineages. Transactions American Fisheries Society 128: 58-87.

Seeb, L.W., P. A. Crane, C. M. Kondzela, R. L. Wilmot, S. Urawa, N. V. Varnavskaya, and J. E. Seeb. 2004a. Migration of Pacific Rim chum salmon on the high seas: insights from genetic data. Environmental Biology of Fishes 69: 21-36.

Seeb, L. W., C. T. Smith, W. D. Templin, R. L. Wilmot, and, J. E. Seeb. 2004b. Development of a Pacific Rim baseline for chum salmon based on single nucleotide polymorphism markers (SNPs). (NPAFC Doc. 824) 10 pages. Alaska Department of Fish and Game, 333 Raspberry Road, Anchorage, AK, USA 99518.

Smith, C. T., J. Baker, L. Park, L. W. Seeb, C. Elfstrom, S. Abe, and J. E. Seeb. 2005. Characterization of 13 single nucleotide polymorphism markers for chum salmon. Molecular Ecology Notes 5: 259-262.

Smith, C. T., C. M. Elfstrom, L. W. Seeb, and J. E. Seeb. In press. Use of sequence data from rainbow trout and Atlantic salmon for SNP detection in Pacific salmon. Molecular Ecology.

Smith, C. T., J. E. Seeb, and L. W. Seeb. 2004. Development of 5'-nuclease reations for high-throughput SNP genotyping in salmon. N. Pac. Anadr. Fish Comm. Tech. Rep. 5: 45-46.

Urawa, S., T. Azumaya, P. A. Crane, and L. W. Seeb. 2004. Origin and distribution of chum salmon in the Bering Sea during the early fall of 2002: estimates by allozyme analysis. (NPAFC Doc. 794) 11 p. National Salmon Resources Center, Toyohira-ku, Sapporo 062-0922, Japan.

Urawa, S., M. Kawana, G. Anma, Y. Kamei., T. Shoji, M. Fukuwaka, K. M. Munk, K. W. Myers, and E. V. Farley, Jr. 2000. Geographic origin of high-seas chum salmon determined by genetic and thermal otolith markers. North Pacific Anadromous Fish Commission Bulletin 2: 283-290.

Werner, F. A. O., G. Durstewitz, F. A. Habermann, G. Thaller, W. Kramer, S, Kollers, J. Buitkamp, M. Georges, G. Brem, J. Mosner, and R. Fries. 2004. Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. Animal Genetics 35: 44-49.

Wilmot, R. L., C. M. Kondzela, C. M. Guthrie, and M. S. Masuda. 1998. Genetic stock identification of chum salmon harvested incidentally in the 1994 and 1995 Bering Sea trawl fishery. North Pacific Anadromous Fish Commission Bulletin 1: 285-299.

Wilmot, R. L., C. M. Kondzela, C. M. Guthrie III, A. Moles, J. J. Pella, and M. Masuda. 2000. Origins of salmon seized from the F/V Arctic Wind. (NPAFC Doc. 471) 18 pp. Auke Bay Fisheries Laboratory, Alaska Fisheries Science Center, NMFS, NOAA, 11305 Glacier Highway, Juneau, AK 99801-8626.

Winans, G. A., P. B. Aebersold, Y. Ishida, and S. Urawa. 1998. Genetic stock identification of chum salmon in high-seas test fisheries in the western North Pacific Ocean and Bering Sea. North Pacific Anadromous Fish Commission Bulletin 1: 220-226.

Table 1. Number of chum salmon populations analyzed for each area within regions of the Pacific Rim.

Region Number	Reporting Region	Area	No. Pop
1	Japan Honshu	Japan Sea Coast of Honshu	2
2	Japan Hokkaido	Okhotsk Sea Coast of Hokkaido	1
2		Pacific Coast Hokkaido	1
3	Russia	Amur River	1
3		Eastern Kamchatka	1
3		Western Kamchatka	2
4	Western Alaska Summer	Kotzebue Sound	3
4		Norton Sound-Port Clarence	2
4		Norton Sound-Subdistricts 1, 2, 3, 6	7
4		Yukon River Summer Lower	4
4		Yukon River Summer Middle	3
4		Kuskokwim Summer	9
4		Bristol Bay Egegik	5
5	Yukon River Fall	Yukon River Fall US	4
5		Yukon River Fall Canada	4
6	Kuskokwim Fall	Kuskokwim Fall	2
7	North Alaska Peninsula	North Alaska Peninsula	2
8	South Alaska Peninsula	South Alaska Peninsula	2
9	Kodiak Island	Kodiak Island	3
10	Susitna River	Susitna River	1
11	Southcentral Alaska	Prince William Sound	2
12	Northern Southeast	Northern Southeast	1
13	Southern Southeast	Prince of Wales Island	1
13		Southeast Alaska Mainland	2
14	Washington Fall	Hood Canal Fall	3
14		Southern Puget Sound	1
14		Strait of Juan de Fuca	1
15	Washington Summer	Hood Canal Summer	3

Table 2. Mitochondrial (mtDNA) and nuclear single nucleotide polymorphisms assayed in chum salmon. Sources describing each SNP giving conditions for genotyping via the 5'-nuclease reaction are given.

Locus	Source
mtDNA SNPs	
Oke_Cr231	Sato et al. 2001; Smith et al. 2005
Oke_Cr30	Sato et al. 2001; Smith et al. 2005
Oke_Cr386	Sato et al. 2001; Smith et al. 2005
Oke_Cr42	Sato et al. 2001
Oke_Cr96	Sato et al. 2001
Oke_ND3-69	Smith et al. 2005
nuclear SNPs	
Oke_arf-31	Smith et al. In Press
Oke_BAMBI-116	Smith et al. In Press
Oke_CKS_2-389	Smith et al. 2005
Oke_copa-211	Smith et al. In Press
Oke_DM20-548	Smith et al. 2005
Oke_eif4ebp2-64	Smith et al. In Press
Oke_GHII-2943	Unpublished
Oke_GHII-3129	Unpublished
Oke_GnRH_3-373	Smith et al. 2005
Oke_GnRH-527	Smith et al. 2005
Oke_GPDH-191	Smith et al. In Press
Oke_HGFA-319	Smith et al. In Press
Oke_hsc71-199	Smith et al. In Press
Oke_il-1racp-67	Smith et al. In Press
Oke_IL8r2-406	Smith et al. 2005
Oke_IL8r-272	Smith et al. 2005
Oke_Moesin-160	Smith et al. In Press
Oke_ras1-426	Unpublished
Oke_RFC2-618	Smith et al. In Press
Oke_RH1op-245	Smith et al. In Press
Oke_SClkF2R2-239	Smith et al. In Press
Oke_serpin-140	Smith et al. In Press
Oke_Tsha1-196	Smith et al. In Press
Oke_u1-519	Smith et al. 2005
Oke_u202-131	Smith et al. In Press
Oke_u212-87	Smith et al. In Press
Oke_u216-222	Smith et al. In Press
Oke_u217-172	Smith et al. In Press
Oke_u200-385	Smith et al. In Press
Oke_Zp3b-314	Smith et al. In Press

Table 3. Results of simulation analyses from two mixtures of Pacific Rim chum salmon. Expected, estimated, and standard deviations are given for both the specific population and reporting region.

	Population			Reporting Region		
Population/Reporting Region	Expected	Estimate	Std. Dev.	Expected	Estimate	Std.Dev.
Mixture 1						
Japan Honshu				0.200	0.196	0.034
Gakko/Japan Honshu	0.100	0.096	0.030			
Sasauchi R./Japan Honshu	0.100	0.100	0.032			
Japan Hokkaido				0.200	0.202	0.035
Shari R./Japan Hokkaido	0.100	0.097	0.028			
Tokachi R./ Japan Hokkaido	0.100	0.105	0.030			
Russia				0.400	0.397	0.027
Amur R/ Russia	0.100	0.091	0.021			
Bistraya River /Russia	0.100	0.093	0.031			
Bolshaya R./Russia	0.100	0.111	0.035			
Palana R/Russia	0.100	0.101	0.026			
Meshik R./W. Alaska Summer	0.100	0.071	0.021	0.100	0.097	0.018
Frosty/North Peninsula	0.100	0.087	0.018	0.100	0.092	0.018
Mixture 2						
Chena/W. Alaska Summer	0.100	0.069	0.036	0.100	0.122	0.037
Yukon Fall				0.400	0.382	0.039
Sheenjek/Yukon Fall	0.200	0.149	0.047			
Donjek/Yukon Fall	0.200	0.177	0.032			
Lawrence/North Peninsula	0.500	0.437	0.037	0.500	0.458	0.032

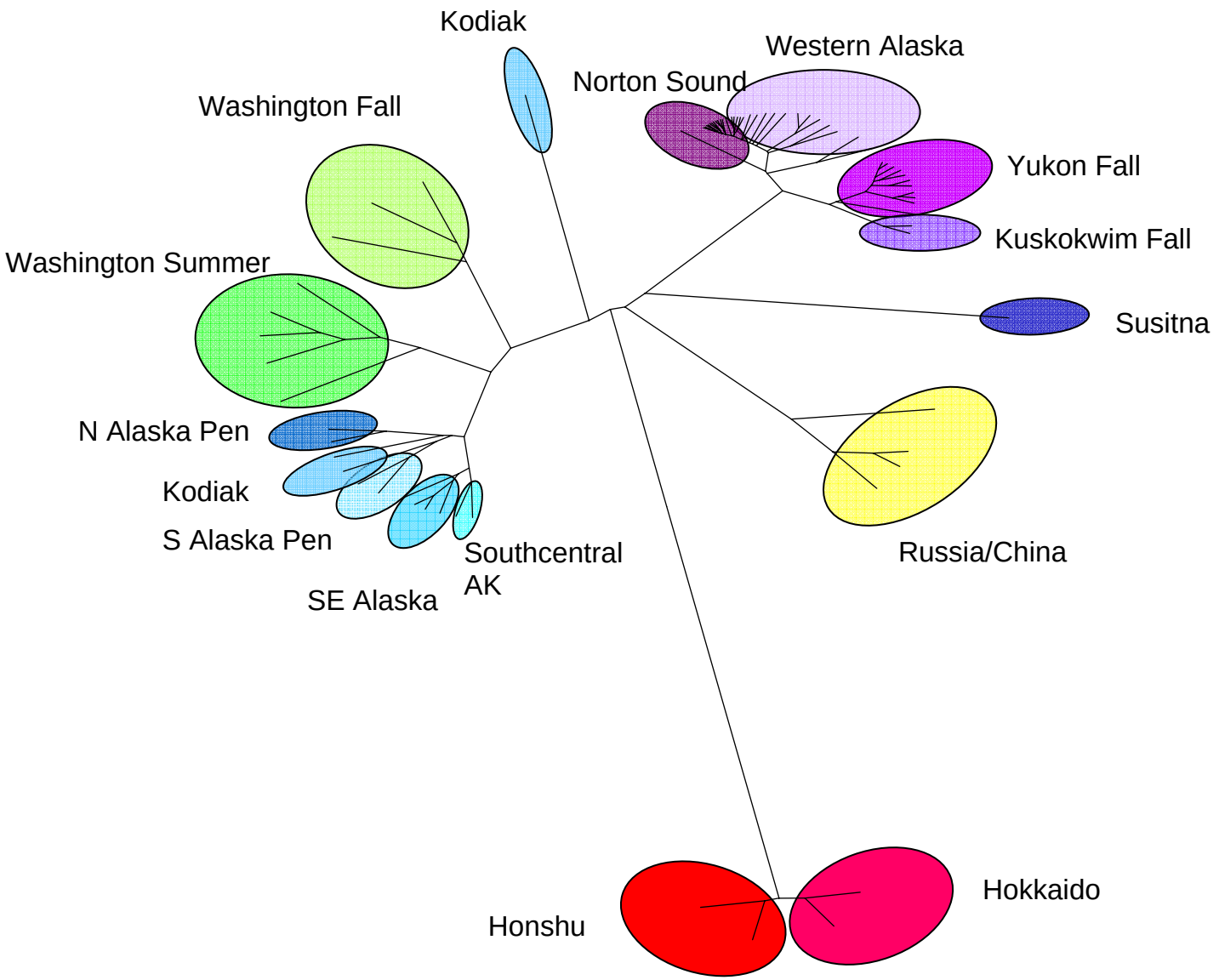

Figure 1. Unrooted UPGMA tree based on Cavalli-Sforza and Edwards (1967) distances.

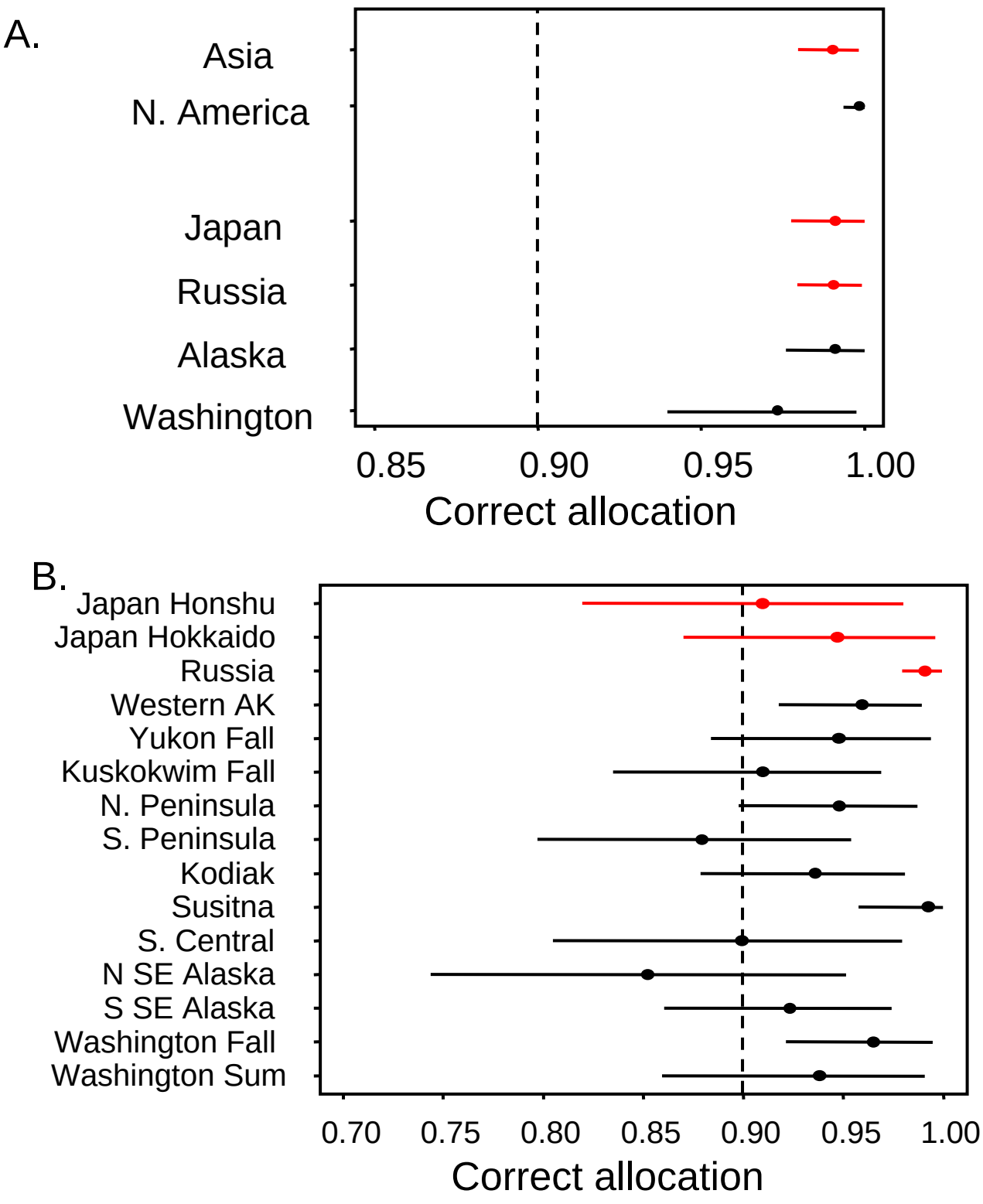

Figure 2. Results from 100% simulations. A. Continents and major regions. B. Reporting regions.