MIGRATION

A complex phenotype in salmon controlled by a simple change in migratory timing

Neil F. Thompson[1,2,3]*, Eric C. Anderson[2,3,4]*†, Anthony J. Clemento[2,3], Matthew A. Campbell[2,3,5], Devon E. Pearse[3,5], James W. Hearsey[6], Andrew P. Kinziger[6], John Carlos Garza[1,2,3]*†

Differentiation between ecotypes is usually presumed to be complex and polygenic. Seasonal patterns of life history in salmon are used to categorize them into ecotypes, which are often considered "distinct" animals. Using whole-genome sequencing and tribal fishery sampling of Chinook salmon, we show that a single, small genomic region is nearly perfectly associated with spawning migration timing but not with adiposity or sexual maturity, traits long perceived as central to salmon ecotypes. Distinct migration timing does not prevent interbreeding between ecotypes, which are the result of a simple, ancient polymorphism segregating within a diverse population. Our finding that a complex migratory phenotype results from a single gene region will facilitate conservation and restoration of this iconic fish.

Biodiversity exists on a continuum, but human perception of biological differences is often distorted by the cognitive benefits of categorization (*1*–*4*). Conspicuous phenotypic differences may serve as the basis for such categorization, but there can be a mismatch between the extent of perceived distinction and the biological basis that underpins it. In some cases, substantial phenotypic differences may be entirely environmental, with no genetic differences between individuals with alternative phenotypes (*5*, *6*).

The seasonal timing of animal migrations is a particularly conspicuous trait, with indigenous peoples around the world using these migrations as the basis for customs related to subsistence and culture. Many human societies have historically depended upon the fishes in the family Salmonidae, which includes salmon, trout, and charr, as a central component of their subsistence economies. Among the most well-known animal migrations are those made by anadromous salmonids, which travel up to 3000 km from the ocean to spawn in natal streams (*7*). In Chinook salmon (*Oncorhynchus tshawytscha*), the largest of the anadromous salmonids, multiple ecotypes exist and are characterized primarily by migration timing (*8*). Early-migrating fish are generally termed winter-run and spring-run ecotypes, whereas late-migrating fish include the fall-run and late-fall–run ecotypes (*7*). Early-migrating salmon return to fresh water up to 6 months before late-migrating fish and hold during summer in cold-water pools until spawning. In addition, early-migrating fish have a suite of other prominent phenotypic differences, including smaller size, higher fat content, earlier spawning time, and the capacity to use different habitats (*9*–*11*), often at higher elevation, where they play a distinctive ecological role, conveying marine-derived nutrients and altering food webs. They are prized by fishers for their higher fat content and presence in fresh water when other salmon are not available. These traits have led many to regard the spring run as an entirely distinct type, or species, from the fall run (*12*). This sentiment mirrors the idea that migration timing is a "magic" trait that causes assortative mating, and potentially sympatric reproductive isolation, and which is also under disruptive selection (*13*).

Two broad themes have emerged in the study of seasonal animal migrations. First, successfully adopting new migration patterns requires a "syndrome" of physiological, morphological, and life-history traits to mediate the trade-offs involved with changes in migration (*14*, *15*). Second, by exposing animals to new habitats, migration can promote reproductive isolation and speciation (*16*). Genomic studies accord with these two ideas. In birds, migration-linked genes have been implicated in divergence between subspecies with distinct migratory directions (*17*), and migratory forms of the same species can display different genomic and transcriptional patterns (*18*–*20*). In fishes, genomic variation associated with distinct migratory behaviors has been shown to be maintained in chromosomal inversions spanning hundreds of genes (*21*, *22*).

Recently, a region on Chinook salmon chromosome 28 that contains two protein-coding genes, *GREB1L* and *ROCK1*, was found to be consistently associated with run-timing ecotypes in multiple drainages of North America (*23*–*25*). It has been suggested that this association is due to genetic adaptation for increased fat storage to offset early migrators' extended lack of feeding in fresh water (*23*). This migration phenotype has also been linked to reproductive maturity, with the adoption of "premature" and "mature" migrator nomenclature (*23*). As changes in migration have precipitated evolution in other key traits, it has been suggested that this trait variation is leading to reproductive isolation sufficiently profound that early- and late-migrating salmon should be categorized and managed as separate species (*26*).

We used whole-genome sequences, targeted genotyping assays, and individual-specific phenotype information from an indigenous fishery to elucidate the genetic architecture of ecotypic differentiation in Chinook salmon and determine the origin of phenotypic variation between ecotypes in these iconic fish. We focus on the Klamath River, where massive restoration efforts, including the removal of four mainstem dams, portend a renaissance for anadromous fishes. We also evaluate the role of genomic variation in ecotypic differentiation in the Sacramento River basin (Fig. 1), which harbors the greatest salmon run–type diversity known.

We resequenced the whole genomes of 160 fish from all ecotypes of Chinook salmon in the Klamath and Sacramento river basins (two of the largest rivers tributary to the Eastern Pacific Ocean), including fish from winter-, spring-, fall-, and late-fall–run ecotypes (Fig. 1, table S1, and data S1). Fish in these basins are not sister lineages evolutionarily, despite their geographic proximity (*27*, *28*). Notably, in a combined analysis of these lineages, fixed differences between 64 fall-run and 64 spring-run fish were found only in a single ~140-kb region on chromosome 28 (Fig. 2A). Within this region, a smaller ~30-kb region, the "region of strongest association" (RoSA), included variants that were also fixed in the winter run, an early-migrating ecotype endemic to the Sacramento River (Fig. 2B).

Variation in the RoSA is organized in distinct, reciprocally monophyletic early- (E) and late- (L) migrating haplotype lineages (Fig. 3). Notably, the Sacramento basin possesses two divergent haplogroups within both the E and L lineages, whereas the Klamath basin has only one per lineage. Early-migrating, spring-summer–run Chinook salmon from the Upper Columbia River (Idaho, USA) are highly divergent from all other Chinook salmon (*27*–*29*) and carry a similar, but nonidentical, E-lineage haplotype in the RoSA (figs. S1 and S2) (*25*). This suggests that the E-lineage haplotype may be shared by all early-migrating Chinook salmon. We aligned the RoSA with the homologous region in its closest relative, coho salmon (*O. kisutch*), and estimate that the divergence between the E and L haplogroups is ~6% of that between coho and Chinook

[1]Department of Ocean Sciences, University of California, Santa Cruz, Santa Cruz, CA 95064, USA. [2]University of California, Santa Cruz, Institute of Marine Sciences, Santa Cruz, CA 95064, USA. [3]Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, Santa Cruz, CA 95060, USA. [4]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA. [5]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA. [6]Department of Fisheries Biology, Humboldt State University, Arcata, CA 95521, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: eric.anderson@noaa.gov (E.C.A.); carlos.garza@noaa.gov (J.C.G.)

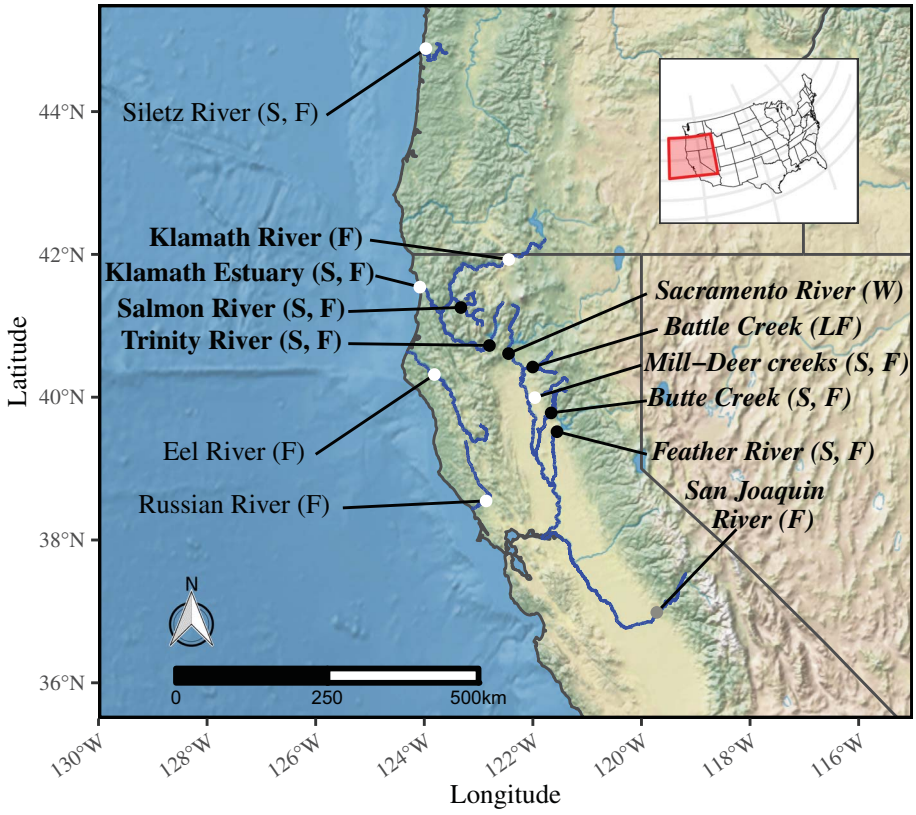

Fig. 1. Waterways from which Chinook salmon were sampled. Ecotypes are represented by letters (W, winter; S, spring; F, fall; LF = late-fall). Collections in bold are from the Klamath River basin, and names in italic-bold are from the Sacramento River basin. White circles indicate amplicon data; gray circles, whole-genome sequence data; and black circles, both data types.

salmon, indicating that the split occurred relatively recently within the Chinook salmon lineage (fig. S3).

Two of the fixed differences in the RoSA are nonsynonymous, single-nucleotide polymorphisms (SNPs) within *GREB1L* and show a near-perfect association between early- and late-migrating genotypes (table S2). *GREB1L* is a central regulator of vertebrate development, specifically affecting renal, gonadal, and inner ear organ systems (*30*, *31*). Additionally, there are structural changes—short duplications between *GREB1L* and *ROCK1* (fig. S4) that might affect gene regulation and are strongly associated with distinct haplogroups (figs. S5 and S6). The structural differences might interact with the nonsynonymous SNPs or other variants to influence the phenotype. Evident within the RoSA is a block-like structure of allelic variation, high conservation, and linkage disequilibrium (Fig. 2B), but evaluation of sequences from this region did not yield evidence of a large inversion. The high conservation in this region may be attributable to its function as a small supergene, with multiple coadapted variants contributing to the phenotypic effects and large fitness differences associated with recombinant haplotypes.

We evaluated the effect of RoSA genotype on the phenotype of 502 adult Chinook salmon harvested by the Yurok Tribe in the Klamath River Estuary (table S1). We assayed eight SNPs (table S3 and data S2) that perfectly tag the E and L haplogroups (fig. S2) and compared the RoSA genotype to the phenotypic traits of freshwater entry date, fat content, and reproductive maturity. We found that the RoSA genotype predicted 85% of the variance in freshwater entry timing (table S4), with nearly disjunct freshwater entry windows among EE and LL genotypes (Fig. 4A). Heterozygotes enter with intermediate timing, on average, but overlap completely with the timing of homozygous fish. Thus, the RoSA genotype directly influences individual freshwater entry timing and, by extension, the initiation of spawning migration during the ocean phase. It may be that the RoSA influences individual response to photoperiod, as this is used by most organisms, including salmon (*32*), to detect seasonal changes, but other physiological processes are likely also involved.

By contrast, the RoSA genotype had no significant effect on maturation status or adiposity at freshwater entry [Fig. 4, B and C; likelihood ratio tests (LRT), table S5A], after accounting for significant effects of sampling date, sex, and collection year (LRT, table S5, B and C). The strongest effect on maturation status and adiposity was due to sampling date, with partial correlations of 0.53 and −0.26, respectively. Notably, all estuary-sampled fish had maturation-status values well below those of spawning fish (fig. S7), contrasting with the recent categorization of fall-run salmon as "mature migrators" (*7*, *23*, *24*).

The RoSA genotype explained 67% of the variance in spawn timing at Trinity River Hatchery (table S6), where EE and LL individuals have a mean difference in spawn timing of 45 days (fig. S8). However, because the RoSA genotype does not affect maturation status at freshwater entry, this difference is likely a consequence of different freshwater residence durations, with warmer freshwater temperatures (*33*), relative to the ocean (fig. S9), accelerating maturation in early-entering fish during summer months. Water temperature is known to influence the rate and onset of final maturation in fishes, including salmonids (*34*, *35*). Thus, although the reproductive physiology of spring- and fall-run fish is indeed distinct, validating indigenous knowledge, the RoSA directly influences freshwater entry timing, with other aspects of the spring-run "syndrome" (adiposity and maturity) attributable to differences in sampling date and environmental factors experienced as a consequence of differences in freshwater entry timing. We thus provide a counter-example to the widely held notion that differences in lipid metabolism and storage are central to migratory biology in vertebrates.

We next characterized the extent to which spawning behavior contributes to assortative mating and distinct populations in natural areas where both ecotypes persist, by evaluating differences in mating location and timing by RoSA genotype in the Salmon River (Klamath River tributary, which is mostly wilderness). The distribution of RoSA genotypes from 183 post-spawning carcasses indicates substantial overlap and opportunity for interbreeding among RoSA genotypes (figs. S10 and S11). This finding is bolstered by a large number of heterozygotes (table S7) and RoSA genotype frequencies close to those expected with random mating (fig. S12). As such, matings between heterozygotes must occur with some frequency, leading to full-sibling families that express both early- and late-migrating phenotypes.

We expanded our RoSA-marker survey to populations ranging from Coastal Oregon to the Sacramento basin (Fig. 1 and table S1). Heterozygous (EL) fish were widespread where early-migrating fish occur and suitable habitat for them exists (table S8 and data S3), and not only where hatchery propagation, or other anthropogenic influences, maintain them. However, we investigated whether spring-run and

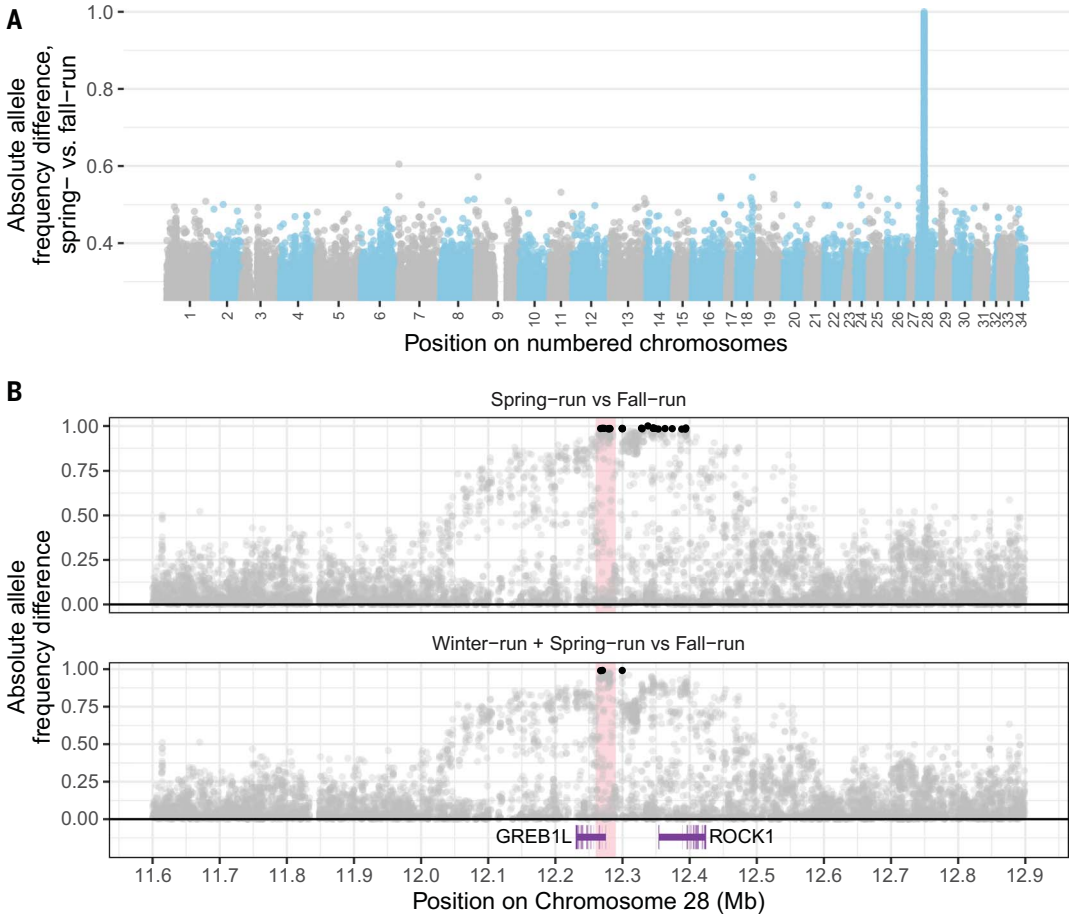

Fig. 2. Allele frequency differences between ecotypes of Chinook salmon. (**A**) Absolute value of allele frequency differences between 64 spring-run and 64 fall-run Chinook salmon. Each point is a genomic variant (differences <0.25 not shown). Colors alternate by chromosome. (**B**) Absolute value of allele frequency differences in a 1.3-Mb region around the peak in (A). Black points indicate nearly fixed (>0.98) differences. Vertical pink rectangle shows the RoSA. Violet horizontal bars at bottom show two nearby genes and exons within them (thin vertical bars).

fall-run fish were historically reproductively isolated via a simulation analysis of recombinant genotypes between previously reported SNPs in the imperfectly associated flanking region (*23*) and the RoSA (figs. S2 and S13 and table S3). We found that there is a long history of interbreeding between spring- and fall-run Chinook salmon in the Klamath River, consistent with earlier work (*36*), and that the high frequencies of recombinants (table S9 and data S4), necessarily generated in heterozygous fish, could not have arisen solely in the ~180 years since large-scale human manipulation of the watershed began (fig. S14 and table S10).

Disruptive selection between salmon ecotypes has been hypothesized to be strong, with heterozygotes particularly vulnerable (*23*, *24*). Our data indicate that the RoSA has a partially dominant or additive inheritance pattern in the Klamath basin. Heterozygotes have migration timing that is skewed toward the early-migrating ecotype but overlaps entirely with the two homozygous classes. Moreover, we find that RoSA heterozygotes in the Sacramento basin are phenotypically more similar to late-migrating, than early-migrating, ecotypes for traits related to spawning (table S8), indicating that the dominance relationship of the RoSA may be lineage specific and influenced by modifier loci. Unidentified modifier genes of smaller effect are known to influence within-season variation in migration timing of salmonids (*37*). Ultimately, it is clear that with temporally variable selection, fish with intermediate migration timing (heterozygotes) will periodically have equal or higher fitness than fish migrating earlier or later.

The importance of early-migrating fish to indigenous peoples and ecological functioning cannot be overstated, and early-migrating haplotypes are undoubtedly an important component of diversity in ecosystems and populations where they exist. Early-arriving fish are a critical early-season food source for indigenous peoples, and the early-migrating phenotype allows fish to exploit habitats that are less accessible to late-migrating fish (*10*), providing increased fitness and resilience to salmon populations, as well as important ecological effects, including deposition of marine-derived nutrients and altered food webs. In some cases, such as the upper Columbia River, early- and late-migrating lineages coexist with little gene flow, are not closely related, and evolved separately over long periods of allopatry (*27*–*29*). In general, however, fall- and spring-run Chinook salmon interbreed, and the primary axis of genomic variation is associated with geography. As such, genome-wide adaptation to local environmental conditions is shared by sympatric spring- and fall-run ecotypes (*27*, *29*, *36*), and the gene flow between them provides critical connectivity for the maintenance of genetic diversity and long-term viability, consistent with a more inclusive definition of salmon populations (*38*–*40*).

The finding that the RoSA E-lineage haplogroup is conserved across Chinook salmon lineages and evolutionarily significant units (i.e., "species" under the U.S. Endangered Species Act) is a positive development for conservation, indicating that the most important diversity related to the evolution of ecotypes has not been lost with widespread extirpation of early-migrating salmon. As restoration efforts continue, the reestablishment of early-migrating populations will be facilitated by the evolutionary conservation and exchangeability of the E-lineage haplogroup. In addition, this highlights the importance of maintaining migratory opportunities and viable early-migrating source populations for recolonization.

In the Klamath River, the largest fish restoration project in history, removal of four dams that have blocked access to the upper basin

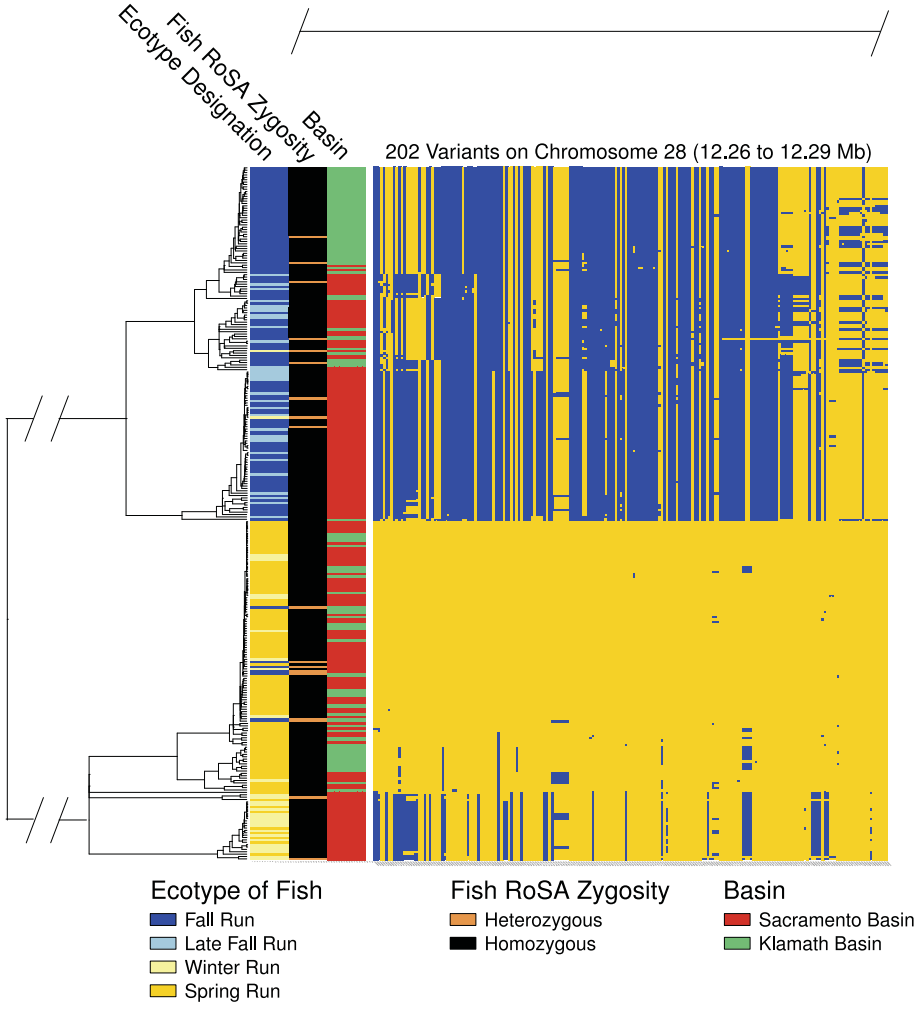

Fig. 3. Two hundred and two biallelic variants (columns) within the RoSA (chromosome 28, 12.26–12.29 Mb). Variants were from 202 haplotypes (rows) derived from whole-genome resequencing of 146 fish, with gold indicating the allele at highest frequency among spring run and blue the allele at lowest frequency among spring run. The local coalescent tree is at left, with the length of portion excised from the internal branches shown at top. Ecotype of the fish is in the first column right of the tree. The RoSA Zygosity column indicates whether the fish carried both an E- and an L-lineage haplotype (heterozygous) in the RoSA.

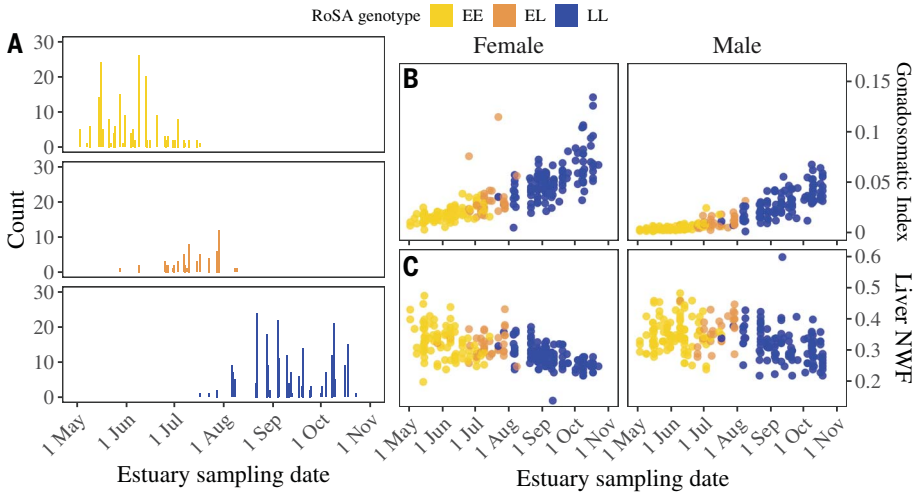

Fig. 4. Migration timing, physiological status, and RoSA genotype of Klamath River Chinook salmon (**A**) Distribution of freshwater entry timing by day and RoSA genotype of 502 Chinook salmon sampled in the Klamath River estuary. (**B**) Maturation status, and (**C**) adiposity of estuary fish by sampling date and color coded by RoSA genotype. Data from 2009 and 2010 are combined.

(~22,600 km^2) since 1912 (*41*) is imminent. In populations where the E haplotype occurs, it is highly likely that an early-run phenotype can emerge given suitable habitat. We found the E haplotype to be absent from adult returns to Iron Gate Dam (table S8), the current terminus of anadromy in the Klamath River, indicating extirpation of the ancestral spring run from the upper Klamath basin. The lack of cool water required for summer survival in the mainstem Klamath River below Iron Gate Dam selects against both EE and EL genotypes during the summer months and is likely responsible for this absence (*33*). However, our analysis of recombinants shows that descendants of the extirpated spring run persist below Iron Gate Dam, likely maintaining much of the genetic variation, outside of the RoSA, that resulted

from local adaptation to the upper Klamath basin. In addition, the E lineage remains abundant in the Klamath River basin (tables S8 and S11) but mostly as a consequence of the Trinity River Hatchery (*36*), which is ~450 river km from the dam removal site. As such, reconstituting a historically accurate, locally adapted, spring-run is possible by migration of E haplotypes into the population currently residing below Iron Gate Dam. Given the strong pattern of equilibrium between migration and genetic drift in the basin (*40*), E haplotypes will arrive in the upper Klamath basin through intrabasin migration, but this could be facilitated more rapidly by human-assisted translocation.

An association between ecotypic designation and a genomic region homologous to the RoSA has also been demonstrated in steelhead trout (*O. mykiss*), with similar ecological importance associated with the early-migrating ecotype (*23*, *42*, *43*). In addition, the three primary genera in the subfamily Salmoninae all have distinct early- and late-migrating forms that co-occur (*7*). We show that a complex migratory phenotype arises from a single gene region, which may facilitate its existence on multiple genetic backgrounds, and suggests that a similar, simple control of migratory traits may exist in other genera. Further research will elucidate how universally such apparent complexity and ecologically important phenotypic variation arises from a simple Mendelian polymorphism.

REFERENCES AND NOTES

1. J. Winawer *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7780–7785 (2007).
2. B. Park, M. Rothbart, *J. Pers. Soc. Psychol.* **42**, 1051–1068 (1982).
3. D. Roberson, J. R. Hanley, *Curr. Biol.* **17**, R605–R607 (2007).
4. J. W. Howard, M. Rothbart, *J. Pers. Soc. Psychol.* **38**, 301–310 (1980).
5. A. D. Bradshaw, in *Advances in Genetics* (Elsevier, 1965), vol. 13, pp. 115–155.
6. K. Gotthard, S. Nylin, *Oikos* **74**, 3–17 (1995).
7. T. P. Quinn, P. McGinnity, T. E. Reed, *Can. J. Fish. Aquat. Sci.* **73**, 1015–1030 (2016).
8. C. Groot, L. Margolis, *Pacific Salmon Life Histories* (UBC Press, 1991).
9. J. O. Snyder, *Salmon of the Klamath River, California* (California State Printing Office, 1931).
10. P. B. Moyle, *Inland Fishes of California: Revised and Expanded* (Univ. of California Press, 2002).
11. J. W. Hearsey, A. P. Kinziger, *Environ. Biol. Fishes* **98**, 413–423 (2015).
12. K. Langin, *Science* **360**, 590–592 (2018).
13. R. S. Taylor, V. L. Friesen, *Mol. Ecol.* **26**, 3330–3342 (2017).
14. H. Dingle, *J. Ornithol.* **147**, 212–220 (2006).
15. T. Alerstam, A. Hedenström, S. Åkesson, *Oikos* **103**, 247–260 (2003).
16. S. P. Turbek, E. S. C. Scordato, R. J. Safran, *Trends Ecol. Evol.* **33**, 164–175 (2018).
17. K. Ruegg, E. C. Anderson, J. Boone, J. Pouls, T. B. Smith, *Mol. Ecol.* **23**, 4757–4769 (2014).
18. A. M. Fudickar *et al*., *Biol. Lett.* **12**, 20160069 (2016).
19. R. A. Johnston, K. L. Paxton, F. R. Moore, R. K. Wayne, T. B. Smith, *Mol. Ecol.* **25**, 5680–5691 (2016).
20. M. Lundberg *et al*., *Evol. Lett.* **1**, 155–168 (2017).
21. T. Kess *et al*., *Sci. Adv.* **5**, eaav2461 (2019).
22. D. E. Pearse *et al*., *Nat. Ecol. Evol.* **3**, 1731–1742 (2019).
23. D. J. Prince *et al*., *Sci. Adv.* **3**, e1603198 (2017).
24. T. Q. Thompson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 177–186 (2019).
25. S. R. Narum, A. Di Genova, S. J. Micheletti, A. Maass, *Proc. Biol. Sci.* **285**, 20180935 (2018).
26. *Fed. Regist.* **83**, 8410–8414 (2018).
27. P. Moran *et al*., *Can. J. Fish. Aquat. Sci.* **70**, 415–435 (2013).
28. A. J. Clemento, E. D. Crandall, J. C. Garza, E. C. Anderson, *Fish Bull.* **112**, 112–130 (2014).
29. R. S. Waples, D. J. Teel, J. M. Myers, A. R. Marshall, *Evolution* **58**, 386–403 (2004).
30. P. D. Brophy *et al*., *Genetics* **207**, 215–228 (2017).
31. I. Schrauwen *et al*., *Hum. Genet.* **137**, 459–470 (2018).
32. T. P. Quinn, *The Behavior and Ecology of Pacific Salmon and Trout* (Univ. of Washington Press, 2018).
33. C. Z. Romberger, S. Gwozdz, "Performance of water temperature management on the Klamath and Trinity Rivers, 2017" (Arcata Fisheries Data Series DS 2018-59, U.S. Fish and Wildlife Service).
34. N. W. Pankhurst, H. R. King, *J. Fish Biol.* **76**, 69–85 (2010).
35. N. W. Pankhurst, M. J. R. Porter, *Fish Physiol. Biochem.* **28**, 385–389 (2003).
36. A. P. Kinziger, E. J. Loudenslager, D. G. Hankin, E. C. Anderson, J. C. Garza, *N. Am. J. Fish. Manage.* **28**, 1426–1438 (2008).
37. A. Abadía-Cardoso, E. C. Anderson, D. E. Pearse, J. C. Garza, *Mol. Ecol.* **22**, 4733–4746 (2013).
38. D. E. Schindler *et al*., *Nature* **465**, 609–612 (2010).
39. S. M. Carlson, W. H. Satterthwaite, *Can. J. Fish. Aquat. Sci.* **68**, 1579–1589 (2011).
40. A. P. Kinziger, M. Hellmair, D. G. Hankin, J. C. Garza, *Trans. Am. Fish. Soc.* **142**, 1347–1357 (2013).
41. J. B. Hamilton *et al*., *Oreg. Hist. Q.* **117**, 326–377 (2016).
42. J. E. Hess, J. S. Zendt, A. R. Matala, S. R. Narum, *Proc. Biol. Sci.* **283**, 20153064 (2016).
43. S. J. Micheletti, A. R. Matala, A. P. Matala, S. R. Narum, *Mol. Ecol.* **27**, 128–145 (2018).
44. E. C. Anderson *et al.*, 2020-chinook-salmon-migration-timing: Code/data repository for Thompson *et al*., Version v1.2, Zenodo (2020); http://doi.org/10.5281/zenodo.3836196.

ACKNOWLEDGMENTS

We are extremely grateful to the Yurok Tribe for allowing us access to their tribal fishery. We thank the field biologists that sampled fish; C. Davis and M. Banks for Siletz River samples; E. Campbell, C. Columbus, and E. Correa for laboratory aid; R. Corbett-Detig for computational aid; E. Bjorkstedt and C. Edwards for temperature data; E. Palkovacs for helpful comments on the manuscript; and the UCLA IDRE for the Hoffman2 Shared Cluster. **Funding:** CIMEC (NOAA) award NA15OAR4320071; California Sea Grant/Delta Science Fellowship Program award R16AC00026; Hoopa Valley Tribal Fisheries Project N04351. **Author contributions:** N.F.T., E.C.A., A.J.C., M.A.C., D.E.P., A.P.K., and J.C.G. conceptualized the study. A.J.C. and M.A.C. curated molecular and whole-genome data. J.W.H. and A.P.K. curated estuary samples and associated data. N.F.T. and E.C.A. conducted formal analysis of the data. N.F.T., E.C.A., and J.C.G. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All genome sequence data are available from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra/PRJNA667732). All other data and analysis code are available in GitHub, archived at Zenodo (*44*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6516/609/suppl/DC1
Materials and Methods
Figs. S1 to S14
Tables S1 to S11
Data S1 to S5
References (*45*–*97*)

21 January 2020; accepted 18 August 2020
10.1126/science.aba9059