

PCR, Direct Sequencing, and the Comparative Approach

Thomas D. Kocher

Department of Zoology and Program in Genetics, University of New Hampshire,
Durham, New Hampshire 03824

PCR has relieved much of the experimental drudgery of molecular biology and dramatically improved the sensitivity of many diagnostic procedures. These practical advantages are having a profound impact on the fields of molecular systematics and comparative biology. The reason is that PCR has made possible the analysis of previously intractable small, old, and poorly preserved samples and greatly facilitated the rapid cloning and sequencing of large numbers of samples. DNA sequences allow finer resolution of phylogenies, represent stable characters for taxonomic identification, and provide new insight into the patterns of evolutionary change in genes. This review will focus primarily on the impact of PCR on the field of molecular systematics, but I will also argue for the incorporation of a more comparative approach into molecular biology in general. Sequence data can be a "common currency" which can intellectually link evolutionary and molecular biology, to the advantage of both fields.

New Methods for Obtaining Comparative Data

Traditional methods for obtaining DNA sequences were designed to clone a particular segment of DNA just once. While efficient for that purpose, they were too tedious for the sequence analysis of a large number of individuals. Three innovations were required to allow rapid comparative studies at the level of DNA sequence. First was the development of PCR as a way to clone genes from a large number of samples simultaneously.^(1,2)

Second was the application of consensus primers to amplify homologous sequences from related organisms.⁽³⁾ It had been expected, especially for animal mitochondrial DNA (mtDNA),

that lack of sequence information would prevent application of PCR to most species. However, even limited homology has proven sufficient for the design of amplification primers. PCR can often be used to recover distantly related homologues, even when standard heterologous probes fail, because only short stretches of identity to the primer sequence are necessary for a successful amplification of the target sequence.

Finally, techniques to sequence the double-stranded product of PCR were needed. The intractability of relaxed double-stranded DNAs for sequencing is clearly demonstrated by the multitude of papers purporting universal techniques for sequencing PCR products. While each of these protocols undoubtedly works in the hands of its inventor, none is a panacea. One of the most successful sequencing methods was the production of single-stranded template by asymmetric amplification.⁽⁴⁾ Multiple cycles of sequencing (cycle-sequencing, ref. 5), analogous to single-stranded amplification, especially using labeled Sanger terminators, offers the greatest hope for the future.⁽⁶⁾ When used in conjunction with automated sequencers, the full potential of PCR for comparative studies can be realized.⁽⁷⁾

Ancient DNA

The pioneering work of Higuchi⁽⁸⁾ and Pääbo⁽⁹⁾ suggested that DNA might be recovered from ancient remains. Because these samples usually contain extremely small quantities of highly degraded DNA, recovery of sequence information was extremely tedious and required extraordinary technical competence. PCR has revolutionized the study of these materials and has demonstrated that ancient remains can

provide useful phylogenetic insight. At the same time, the deceptive simplicity of PCR has encouraged optimists to ignore the biochemical realities of ancient samples. Old DNA is typically degraded to short molecules, which are often highly modified.⁽¹⁰⁾ Amplification from these templates is typically limited to a few hundred base pairs, and contamination of both the samples and the PCR reactions by exogenous DNA and PCR products is an ever-present danger that must be monitored with careful controls.

The first demonstration of the utility of PCR for the study of ancient remains was the amplification of mtDNA sequences from 7000-year-old human brains recovered from a peat bog.⁽¹¹⁾ Enough phylogenetic information was recovered from the sequence to determine that this individual contained a unique mitochondrial haplotype. Other demonstrations of the effectiveness of PCR were the sequencing of a 40,000-year-old mammoth⁽¹⁰⁾ and the extinct marsupial wolf.⁽¹²⁾ But the most spectacular claims are the amplification and sequencing of chloroplast genes from 17- to 20-million-year-old *Magnolia* and *Taxodium* fossils.^(13,14) Should these sequences withstand further scrutiny,⁽¹⁵⁾ they represent an extraordinary opportunity for the study of molecular evolution.

These impressive achievements had an important spinoff—the development of methods for recovering DNA, at very low concentration, from a broad range of less than ideal samples. The cloning of sequences from the quagga skin encouraged the examination of dozens of kangaroo rat skins to detect changes in gene frequency over historical time scales.⁽¹⁶⁾ It also prompted study of other dried material, especially herbarium specimens.⁽¹⁷⁾

The rejuvenation of these collections for molecular work is one of the most important achievements of the study of ancient DNA.

Liquid preservation has also received considerable attention. The recovery of sequences from alcohol-preserved tissues is now routine,⁽¹⁸⁾ greatly simplifying tissue collection under difficult field conditions. Even samples preserved in anhydrous acid (Carnoy's) are readily accessible.⁽¹⁹⁾ Widespread and reliable use of formalin-fixed material, however, still presents difficulties. Only a few studies have successfully amplified DNA from formalin-preserved museum specimens,⁽²⁰⁾ and it appears likely that the exact conditions of formalin fixation are important. The high degree of cross-linking and other damage to DNA in formalin limits the length of sequences that can be amplified. It would be useful to examine alternative fixation techniques which are compatible with both molecular and morphological studies.

The ability to use PCR to amplify DNA from historical samples allows many new perspectives on the evolution and function of organisms. Just as frozen viral samples offer an opportunity to recover historical information about the structure and function of influenza viruses,⁽²¹⁾ museum specimens offer an opportunity to examine changes in allele frequency and gene structure over evolutionary time. A particularly interesting example is the demonstration that Lyme disease spirochaetes can be amplified from 100-year-old ticks, indicating that the increasing prevalence of this disease is probably not due to the introduction of a new pathogen, but to local increases in abundance of spirochaetes already present in the environment.⁽²²⁾

Intraspecific Phylogenies

PCR allows the study of large numbers of samples and opens for the first time the possibility of describing population variation at the level of DNA sequence. The advantages are both a greater sensitivity (all changes can be detected) and new modes of data analysis (a historical perspective on population variants). Most of these studies have focussed on haploid genomes to avoid complications inherent when two al-

leles are amplified simultaneously from a diploid genome. Studies of animal mtDNA have been particularly abundant, in part because of the long history of comparative studies of mtDNA. In most vertebrates, mtDNA evolves rapidly and is inherited maternally, two properties that are useful for examining the historical relationships among organisms.

The most well-known of these studies have focussed on humans.⁽²³⁾ Despite recent controversy over the particular conclusion of an African origin for our species,⁽²⁴⁾ PCR and direct sequencing are providing a detailed look at the geographic structure of human mitochondrial DNA variation. Many of the studies are made possible by the simplification of sample collection through the use of single, alcohol-preserved hairs rather than blood.⁽²⁵⁾ Furthermore, the data obtained are of much finer resolution than previous studies based on RFLPs because sequencing can be directed at the most variable parts of the mitochondrial genome. This higher resolution was demonstrated by work on European populations⁽²⁶⁾ in which 88 distinct mitochondrial types were identified among 117 individuals. Sequence data may eventually allow statistically significant outgroup rooting of the human mtDNA tree to chimps,⁽²⁷⁾ which has been impossible with restriction site data. These advances are not limited to the well-studied human mtDNA. The population structure of many other species, including fish⁽²⁸⁾ and sea urchins,⁽²⁹⁾ is being examined in this way.

Invisible Organisms

Traditional biochemical methods perpetuated a bias toward the study of large or easily cultured organisms. PCR has greatly expanded the range of taxa that can be studied, with sometimes surprising results. For instance, it has long been known that many naturally occurring bacteria resist laboratory culture. As a result, studies of bacterial communities are seriously deficient in that they recognize perhaps only one in five bacterial species present.⁽³⁰⁾ PCR has been used to assess the diversity of marine bacterial communities by amplification of a portion of the 16S rRNA gene from seawater. The sequences de-

rived from this sample identified at least two new groups of bacteria, including one divergent enough to be recognized as a unique phylum.⁽³¹⁾ Many other small marine species cannot be cultured, and have been difficult to study with traditional molecular methods.⁽³²⁾ PCR is ideally suited to the analysis of symbionts, which typically cannot be cultured outside of their hosts. The flexibility of zooxanthellae-coral associations over evolutionary time⁽³³⁾ has recently been demonstrated by a combination of amplification and sequencing.

Identification of small and morphologically simple organisms is often difficult. Two examples demonstrate the opportunities offered by PCR amplification of specific sequences. The first is the identification of fungi in mycorrhizal associations.⁽³⁴⁾ Single mycorrhizal rootlets can be amplified and the fungal strain identified on the basis of sequence differences. The morphological similarity of the larval forms of many marine invertebrates has hindered ecological work, but amplification and sequencing of mtDNA can be used to identify larval sea cucumbers.⁽³⁵⁾ Species identifications based on DNA sequence differences will allow new insight into life cycles and population structure of marine species in the future.

Taxonomic Database?

As funding for curatorial positions declines in museums around the world, it has become difficult to find qualified individuals to identify and classify organisms. For many groups, even qualified taxonomists cannot find characters for the reliable identification of all life stages of a species. PCR makes it easy to use DNA sequences as a molecular tag for species identity. Why not make it easy for anyone to identify a species, or find out what the close relatives are? The informational complexity of even a short stretch of DNA is sufficient to distinguish a large number of species. Consensus primers applicable to broad groups of organisms could be used to amplify a homologous region from each species. The internal sequence of the fragment can be matched to a database of previously reported sequences and the closest matches recovered. Techniques for

doing this are already well established in existing sequence databases. The specific region to be included in the database could be decided by nomenclature committees, in consultation with students of molecular evolution. For animals, a segment of a mitochondrial rRNA gene might provide sufficient resolution. Sequence similarity would provide criteria to evaluate confidence in the identification, and might provide preliminary information for the construction of phylogenies. An analogous situation is the adoption of the "sequence-tagged sites" approach to mapping the human genome.⁽³⁶⁾ Rather than maintaining permanent collections of human clones that can be used to identify polymorphisms in the human population, gene mappers have begun publishing short regions of DNA sequence to identify these loci. This allows investigators around the world to synthesize PCR primers to amplify a particular locus specifically, without the physical distribution of the original clones. The success of the GenBank/EMBL database, whose usefulness increases daily with the addition of new sequence, provides confidence that this approach is tractable. While a sequence database would in no way replace the knowledge held by the qualified taxonomist, it may represent a rapid and flexible way to organize incompletely studied groups.

Phylogenetic Reconstruction

All methods for reconstructing phylogeny from molecular sequences share a common assumption: that the accumulation of substitutions occurs with some regularity over time and among lineages. A number of indirect methods for measuring genetic variation and divergence in natural populations have been popular⁽³⁷⁾; however, techniques such as DNA-DNA hybridization, microcomplement fixation, allozyme electrophoresis, and RFLP analysis, while readily accessible and appropriate for multiple comparisons, have serious deficiencies. First, they do not reveal all of the structural variation in the sequence being compared, since a single protein electromorph may contain several different allelic proteins.⁽³⁸⁾ Second, these methods do not necessarily identify homologous characters. For instance, the loss of a

restriction site can arise from substitutions at any nucleotide within the recognition sequence. Identification of homologous characters is a critically important prerequisite to phylogenetic analysis.

However, the most significant fault of these methods is that they obscure the rates and patterns of base substitution. The result has been a continuing lack of precision in the usage of the terms sequence *difference* and sequence *divergence*. Sequence difference is the observed number of differences in a comparison of two sequences. Sequence divergence represents the total number of substitutions that have occurred on each lineage since the divergence from a common ancestor. Among very close relatives, the two values are similar, but with time, multiple substitutions at the same site, changes in base composition, and patterns of selection can affect the apparent rate of change. Molecular divergence is the result of an interplay of mutational and selective processes. Only by a direct examination of DNA sequences can the rates and patterns of base substitution be determined.

Understanding Substitution

A classic example of the need for comparative studies concerns the mechanism of evolutionary change in mtDNA. A comparison of the first two mitochondrial genomes sequenced (human and mouse) showed that 45% of the sequence differences were translations ($C \leftrightarrow T$ or $A \leftrightarrow G$). This closely approximates the pattern expected from random substitution of the four nucleotides. The strong transition bias, which is the characteristic feature of animal mtDNA, became evident only when sequences of closely related primates were compared. These comparisons showed that transitions exceeded transversion by a ratio of at least 10:1.⁽³⁹⁾ Thus, the exclusive study of a few model systems missed an important feature of mtDNA biology, the mechanism of which is still under debate. Similar comparative studies are leading to an understanding of strand bias, the strong skew in base composition between the two strands of mtDNA, which has not been approachable with a bottoms-up approach.⁽⁴⁰⁾

Our increasingly detailed knowl-

edge of the pattern of substitution in mtDNA suggests that further refinements of techniques for phylogenetic reconstruction are needed. For too long, molecular systematists have relied on untested assumptions about the symmetry of the substitution matrix and the equivalence of substitutional probabilities among sites, primarily because it was difficult to examine these assumptions with the indirect measures of sequence divergence available. PCR has provided the opportunity to examine these assumptions in detail, so that more accurate models of the substitution process can be developed. The exact pattern of substitution is likely to vary among genes and among groups of species, and it is critically important that these variations be taken into consideration in the reconstruction of phylogenies.

Renaissance of the Comparative Approach

Comparative studies are becoming increasingly important in the organization and unification of results from the variety of model systems currently used by molecular biologists. Functional studies are identifying similarity in the molecular mechanisms among even distant relatives. Underlying this similarity are genes derived from a common ancestor. The identification of homologous genes among distantly related species is critically important to the linkage of the diverse model systems currently in use. Phylogenetic variants are also a valuable source of comparative information, both for determining the structure of the molecule⁽⁴¹⁾ and for insight into the selective forces acting on its function. For example, the analysis of the HLA-DQ α locus in humans and primates has revealed the ancient origin of most alleles, suggesting that they have been maintained within species by overdominant selection.⁽⁴²⁾

It is increasingly apparent that most genes are members of gene families whose members have diverged in both sequence and function. PCR is often the most expedient means to assemble comparative sequence data for analysis of these multigene families. This method has been used to unravel the history of the *Wnt* family of developmental regulators.⁽⁴³⁾ The compara-

tive analysis revealed not only the time of origin of each gene in the family, but also changes in constraint on the proteins as they acquired new functions. Other workers are using the same approach to identify the *Hox* gene complements of a variety of species.⁽⁴⁴⁾

The relatively short homologies necessary for primer binding allow even distantly related homologues to be recovered by PCR.⁽⁴⁵⁾ Comparative studies and linking of model systems may be facilitated by a leap-frogging strategy.⁽⁴⁶⁾ Genes whose function is discovered in a model system (such as *C. elegans*), can be isolated from humans, or other distantly related species, by amplification with primers designed to conserved regions of these genes. Such strategies may be more efficient than the use of heterologous probes and will certainly encourage the inclusion of more taxa in comparative studies.

Conclusion

PCR has allowed many studies to be done faster and easier. Speed has been especially important to the application of DNA sequencing on a comparative scale. But PCR has also allowed some qualitatively different things to be accomplished—the study of ancient DNA, identification of unculturable bacteria, and the cloning of distantly related sequences from small regions of homology. It is just beginning to provide data to assess more critically the rates and patterns of evolutionary change in genetic systems.

The accumulation of comparative sequence data encourages the hope that a reunification of functional and comparative biology will occur.

ACKNOWLEDGMENTS

I thank Ulf Gyllensten for suggesting I learn about PCR long before it became a household word, N. Bowers and A. Klein for comments on the manuscript, and the National Science Foundation for support of my research.

REFERENCES

1. Mullis, K. and F. Faloona. 1987. Specific synthesis of DNA *in vitro* via a polymerase catalysed chain reaction. *Methods Enzymol.* **55**: 335–350.

2. Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H.A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
3. Kocher, T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Pääbo, F.X. Villablanca, and A.C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci.* **86**: 6196–6200.
4. Gyllensten, U.B. and H.A. Erlich. 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQ α locus. *Proc. Natl. Acad. Sci.* **85**: 7652–7656.
5. Murray, V. 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* **17**: 8889.
6. Thomas, W.K. and T.D. Kocher. 1992. Sequencing of PCR-amplified DNAs. *Methods Enzymol.* (in press).
7. Ferl, R.J., C.J. Nairn, J.-X. She, D. Wakeland, and E. Almira. 1991. The application of automated DNA sequence analysis to phylogenetic studies. In *Phylogenetic analysis of DNA sequences* (ed. M.M. Miyamoto and J. Cracraft), pp. 45–58. Oxford University Press, New York.
8. Higuchi R., B. Bowman, M. Freiberger, O.A. Ryder, and A.C. Wilson. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312**: 282–284.
9. Pääbo, S. 1985. Molecular cloning of ancient Egyptian mummy DNA. *Nature* **314**: 644–645.
10. Pääbo, S., R.G. Higuchi, and A.C. Wilson. 1989. Ancient DNA and the polymerase chain reaction. *J. Biol. Chem.* **264**: 9709–9712.
11. Pääbo, S., J.A. Gifford, and A.C. Wilson. 1988. Mitochondrial DNA sequences from a 7,000-year old brain. *Nucleic Acids Res.* **16**: 9775–9787.
12. Thomas, R.H., W. Schaffner, A.C. Wilson, and S. Pääbo. 1989. DNA phylogeny of the extinct marsupial wolf. *Nature* **340**: 465–467.
13. Golenberg, E.M., D.E. Giannasi, M.T. Clegg, C.J. Smiley, M. Durbin, D. Henderson, and G. Zurawski. 1990. Chloroplast DNA sequence from a Miocene *Magnolia* species. *Nature* **344**: 656–658.
14. Soltis, P.S., D.E. Soltis, and C.J. Smiley. 1992. An *rbcL* sequence from a Miocene *Taxodium* (bald cypress). *Proc. Natl. Acad. Sci.* **89**: 449–451.
15. Sidow, A., A.C. Wilson, and S. Pääbo. 1991. Bacterial DNA in *Clarkia* fossils. *Philos. Trans. R. Soc. Lond. B* **333**: 429–433.
16. Thomas, W.K., S. Pääbo, F.X. Villablanca, and A.C. Wilson. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. *J. Mol. Evol.* **31**: 101–112.
17. Bruns, T.D., R. Fogel, and J.W. Taylor. 1990. Amplification and sequencing of DNA from fungal herbarium specimens. *Mycologia* **82**: 175–184.
18. Meyer, A., T.D. Kocher, P. Basasibwaki, and A.C. Wilson. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature* **347**: 550–553.
19. Xiong, B. and T.D. Kocher. 1991. Comparison of mitochondrial DNA sequences of seven morphospecies of black flies (Diptera: Simuliidae). *Genome* **34**: 306–311.
20. Meyer, A. and A.C. Wilson. 1990. Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. *J. Mol. Evol.* **31**: 359–364.
21. Hayashida, H., H. Toh, R. Kikuno, and T. Miyata. 1985. Evolution of influenza virus genes. *Mol. Biol. Evol.* **2**: 289–303.
22. Persing, D.H., S.R. Telford III, P.N. Rys, D.E. Dodge, T.J. White, S.E. Malawista, and A. Spielman. 1977. Detection of *Borrelia burgdorferi* DNA in museum specimens of *Ixodes dammini* ticks. *Science* **249**: 1420–1423.
23. Vigilant, L., M. Stoneking, H. Harpending, K. Hawkes, and A.C. Wilson. 1991. African populations and the evolution of human mitochondrial DNA. *Science* **253**: 1503–1507.
24. Templeton, A.R. 1992. Human origins and analysis of mitochondrial DNA sequences. *Science* **255**: 737.
25. Vigilant, L., R. Pennington, H. Harpending, T.D. Kocher, and A.C. Wilson. 1989. Mitochondrial DNA sequences in single hairs from a Southern African population. *Proc. Natl. Acad. Sci.* **86**: 9350–9354.

26. Di Rienzo, A. and A.C. Wilson. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci.* **88**: 1597–1601.
27. Kocher, T.D. and A.C. Wilson. 1991. Sequence evolution in mitochondrial DNA in humans and chimpanzees: Control region and a protein-coding region. In *Evolution of life: Fossils, molecules, and culture* (ed. S. Osawa and T. Honjo), pp. 391–413. Springer-Verlag, Tokyo.
28. Carr, S.M. and H.D. Marshall. 1991. A direct approach to the measurement of genetic variation in fish populations. *J. Fish Biol.* **39(A)**: 101–107.
29. Palumbi, S.R. and B.D. Kessing. 1991. Population biology of the trans-arctic exchange: mtDNA sequence similarity between Pacific and Atlantic sea urchins. *Evolution* **45**: 1790–1805.
30. Ward, D.M., R. Weller, and M.M. Bateson. 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**: 63–65.
31. Giovannoni, S.J., T.B. Britschgi, C.L. Moyer, and K.G. Field. 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60–63.
32. Medlin, L., H.J. Elwood, S. Stickel, and M.L. Sogin. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**: 491–499.
33. Rowan, R. and D.A. Powers. 1991. A molecular genetic classification of zooxanthellae and the evolution of animal-algal symbioses. *Science* **251**: 1348–1351.
34. Gardes, M., T.J. White, J.A. Fortin, T.D. Bruns, and J.W. Taylor. 1991. Identification of indigenous and introduced symbiotic fungi in ectomycorrhizae by amplification of nuclear and mitochondrial ribosomal DNA. *Can. J. Bot.* **69**: 180–190.
35. Olson, R.R., J.A. Runstadler, and T.D. Kocher. 1991. Whose larvae? *Nature* **351**: 357–358.
36. Olson, M., L. Hood, C. Cantor, and D. Botstein. 1989. A common language for physical mapping of the human genome. *Science* **245**: 1434–1435.
37. Hillis, D.M. and C. Moritz. 1990. *Molecular systematics*. Sinauer, Sunderland, Massachusetts.
38. Coyne, J.A. 1982. Gel electrophoresis and cryptic protein variation. In *Isozymes: Current topics in biological and medical research* (ed. M.C. Rattazzi, J.G. Scandalios, and G.S. Whitt), vol. 6, pp. 1–32. Alan R. Liss, New York.
39. Brown, W.M., E.M. Prager, A. Wang, and A.C. Wilson. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *J. Mol. Evol.* **18**: 225–239.
40. Thomas, W.K. and A.C. Wilson., in preparation.
41. Thornton, J.M., T.P. Flores, D.T. Jones, and M.B. Swindells. 1991. Prediction of progress at last. *Nature* **354**: 105–106.
42. Gyllenstein, U.B. and H.A. Erlich. 1989. Ancient roots for polymorphism at the HLA-DQ α locus in primates. *Proc. Natl. Acad. Sci.* **86**: 9986–9990.
43. Sidow, A. 1992. Diversification of the Wnt gene family on the ancestral lineage of vertebrates. *Proc. Natl. Acad. Sci.* (in press).
44. Kappen, C. and J. Pendleton, personal communication.
45. Sommer, R. and D. Tautz. 1989. Minimal homology requirements for PCR primers. *Nucleic Acids Res.* **17**: 6749.
46. Thomas, W.K. and A.C. Wilson. 1991. Mode and tempo of molecular evolution in the nematode *Caenorhabditis*: Cytochrome oxidase II and calmodulin sequences. *Genetics* **128**: 269–279.



PCR, direct sequencing, and the comparative approach.

T D Kocher

Genome Res. 1992 1: 217-221

Access the most recent version at doi:[10.1101/gr.1.4.217](https://doi.org/10.1101/gr.1.4.217)

References

This article cites 38 articles, 13 of which can be accessed free at:
<http://genome.cshlp.org/content/1/4/217.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
