

Am. J. Hum. Genet. 75:000, 2004

Etruscan Artifacts: Much Ado about Nothing

To the Editor:

Malyarchuk and Rogozin (2004 [in this issue]) and Bandelt (2004 [in this issue]) question the authenticity of the sequences in the study by Vernesi et al. (2004), because these sequences contain substitution motifs that are seldom observed in modern samples. Before answering their criticisms in detail, we would like to make three general points: (1) in the study of ancient DNA, the presence of artifacts is inevitably hard to rule out with certainty, and this also applies to our Etruscan data; however, (2) the Etruscan sequences were determined using a very strict set of standards, and, hence, we are confident that they are rather reliable; and (3) many of Bandelt's speculations are based on sites with a high mutation rate, where substitutions have been observed in various associations. For instance, positions 16126, 16219, 16278, and 16362, found in Etruscan haplotypes that Bandelt (2004 [in this issue]) considers with suspicion, are described as mutational hotspots by Wakeley (1993), Hasegawa et al. (1993), Meyer et al. (1999), and even (three of them) Bandelt et al. (2002). Multiple occurrences of those mutations on different lineages does not prove, or suggest, sequencing artifacts.

To test the reproducibility of the results, in our Etruscan study, we cloned the PCR products and sequenced multiple clones from the same individuals (Handt et al. 1996). *Taq* polymerase errors may lead to the apparent occurrence of mutations unique to a single clone (singletons). These errors are, as completely as possible, eliminated by comparing clones and identifying a consensus sequence. Malyarchuk and Rogozin (2004 [in this issue]) compared variation across modern humans with that observed in the 575 clones derived from 28 Etruscan specimens. They noticed that the mutations in the two data sets overlap only in part and that mutations never observed in modern humans occur in the Etruscan clones. However, they were comparing incomparable quantities. On the one hand, they have the spectrum of mutations carried by modern individuals; on the other, they have a set of mutations comprising both those that were actually in the Etruscans' DNA and those that

probably result from *Taq* errors. There is no reason to expect the two spectra to be similar. As mentioned on page 698 of our article (Vernesi et al. 2004), the misincorporation rate was 0.27% for the Etruscans—that is, less than observed, for instance, by Handt et al. (1994). Therefore, Malyarchuk and Rogozin (2004 [in this issue]) are wrong when they say that the clones of the Etruscan data set show an excess of singletons. On the contrary, if the appropriate comparisons are made, singleton sequences are relatively rare among them. Malyarchuk and Rogozin (2004 [in this issue]) also notice that two mutations (16229 and 16334) independently occurring twice in the Etruscan network are rare or absent in almost 8,000 modern samples. Their observation is correct, but, given the methodologies that we applied, at this stage, we do not think it implies that these mutations are necessarily laboratory artifacts.

Bandelt (2004 [in this issue]) makes five comments, which we list here along with our responses.

- 1. The 16193-16219 motif occurs in association with different substitutions in modern samples and among the Etruscans.** Yes, we already wrote that in the second-to-last paragraph on page 698 (Vernesi et al. 2004). Bandelt notes that the mutation at position 16219 has been observed in two different haplogroups in modern people. We know of at least three (U6a, T*, and H) and, within H, of several different haplotypes with the 16219 mutation (H07, H09, H13, and H16). Meyer et al. (1999) showed that mutation rates at 16219 are fivefold higher than the average mutation rate in this region, and therefore 16219 seems to be a fast-mutating site, which can exist in association with a broad range of other mutations.
- 2. The mutation at 16069 occurs in two individuals, 2V and 11C, and in neither is it associated with a mutation at 16126 or with a cut at -766 *MseI*, which generally characterize haplogroup J.** We published sequence 2V for completeness of information, but we did not consider it in any numerical analyses, because it seemed suspicious to us as well (p. 698). As for sequence 11C, the absence of the restriction cut at 14766 *MseI* was confirmed by sequencing the region around the restriction site. At any rate, a sequence, SCOT0492, with the 16069-16261 motif but with-

out a mutation at 16126, was described by Helgason et al. (2001). Also, 16126 is considered by Bandelt et al. (2002) to be a “speedy site” and, hence, not a site informative for phylogenetic analysis. Finally, Meyer et al. (1999) showed that position 16069 has a higher-than-average mutation rate; hence, whether to consider it as a fast-mutating site seems largely a matter of taste.

3. **We rejected the hypothesis that postmortem changes may have affected the 16069 or 16294 sites.** No, we didn’t. Instead, we wrote (p. 699) that we saw *no compelling reason* to think that postmortem changes had occurred there, because Gilbert et al. (2003) found no instance of changes of that kind in their study. Ruling out postmortem changes completely is probably impossible with the available methodologies.
4. **There are inconsistencies between table 1 and figure 2.** Yes, here Bandelt is right. Haplotype 6AM was erroneously reported in one case with +14766 *MseI*, whereas it has been observed only in association with –14766 *MseI*. Another error is a typo in the table, where 13C should be –14766 (as correctly represented in the network). On the other hand, no triangle should be reconstructed for site 95 (as asked by Bandelt), since the two mutations at this site are different (as correctly reported in the table and on p. 698; a gap in the label in the figure shows that the letter G, indicating a transversion, disappeared in print). He is also right when he points out that there is an error in the network method cited, but the error is in the reference list—namely, Bandelt et al. (1999) is given instead of (1995). However, Bandelt is wrong when he says that the node sizes in figure 2 do not correspond to the haplotype frequencies. Sequences 5AM, 6AM, and 7AC (two occurrences) are represented by big circles, and sequence 14CMT (three occurrences) by a very big circle. The network is a reduced median network (Bandelt et al. 1995, 2000), constructed by hand and checked with the program Netwin 4.1 (Fluxus Engineering Web site). Nucleotide positions were weighted using the list of Hasegawa et al. (1993) and following the weighting procedure of Richards et al. (1998). We further reduced the reticulation in the network by a procedure followed in other mitochondrial studies (Richards et al. 1996; Torroni et al. 2001) (see fig. 1). No conclusions of our article change on the basis of this slightly modified network.
5. **Not all of the strictest criteria for the validation of ancient sequences have been followed, because only three individuals were independently sequenced in two laboratories.** All right, then we shall reformulate our sentence as follows: We are not aware of any studies in which the criteria for the validation of ancient sequences were followed more strictly. See, for

example, Keyser-Tracqui et al. (2003) and Maca-Meyer et al. (2004). Cooper and Poinar (2000) also recommend looking for human DNA in associated faunal material, and we could do that only for one individual, because animal bones were not retrieved from any other burial. We shall try to replicate greater numbers of results in future studies, but it must be understood that it is a very long process and one for which ancient material may not always be easily available.

In brief, we thank Bandelt for detecting some differences between the data reported in table 1 of our article and their graphical representation in the network. However, we suggest that Bandelt read our text more carefully and also consider the lists of fast-mutating sites compiled by authors other than himself. Bandelt is clearly right when he stresses the importance of quality in the data, and we showed that we share the same concern in three ways: (1) by excluding from the analysis 50 of the initial 80 samples, whenever any of eight biochemical tests did not suggest a good probability of obtaining reproducible results; (2) by further excluding two identical sequences from the same burial, to make sure that consanguinity did not bias our results; and (3) by excluding sequence 2V because it showed two sets of mutations previously observed in evolutionarily distant haplogroups, and that could conceivably result from the presence of multiple DNAs in the specimen. On the contrary, the problems Bandelt raises are due to individual substitutions that, he says, are “almost always” observed in different contexts. We think science would proceed very slowly, if at all, if scientists agreed to trust only the data they observe almost always. Mitochondrial data sets contain many homoplasies, and sites that were considered mutational hotspots have changed their status over time. Our understanding of mitochondrial variation is still evolving, and it seems bizarre to dismiss as implausible all the data that do not neatly conform to what we think we already know.

Whereas Malyarchuk, Rogozin, and Bandelt think that the Etruscan sequences are too different from modern ones to be good, Serre et al. (2004) argued that ancient mitochondrial sequences should be considered authentic only if they clearly differ from known modern ones. In other words, to make everybody happy, ancient mtDNA sequences should be at the same time *identical to and different from* the sequences of modern people. The famous lines of Joseph Heller’s *Catch 22* come inevitably to mind: “Orr was crazy and could be grounded. All he had to do was ask; and as soon as he did, he would no longer be crazy and would have to fly more missions.”

With this, we do not mean to deny that problems exist

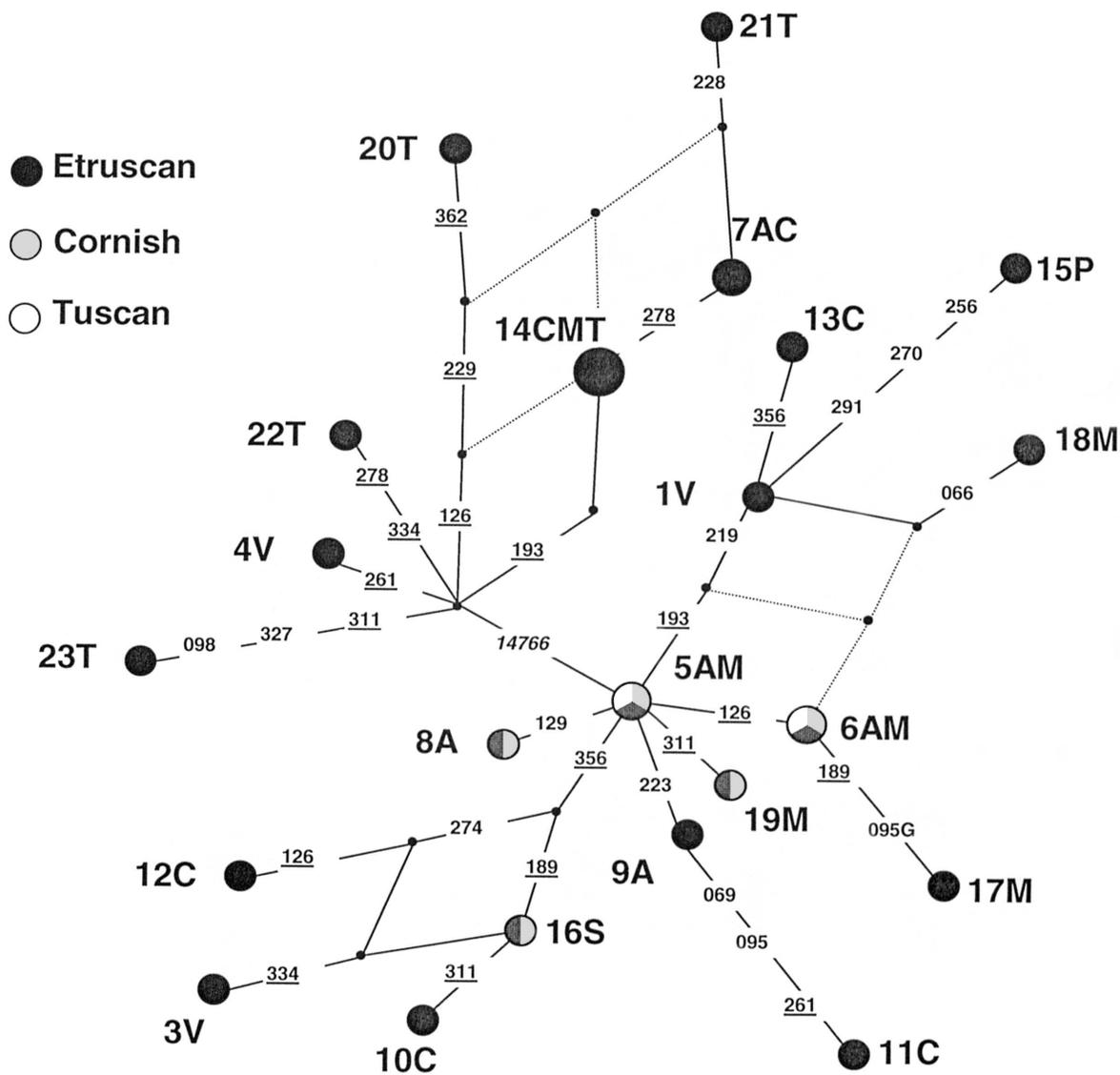


Figure 1 Reduced median network (Bandelt et al. 1995) of sequences identified in Etruscan individuals. The numbers are arbitrary codes; the letters refer to the sites where the sequences have been observed. Haplotype 5AM is the Cambridge Reference Sequence. Nodes are proportional to frequencies. The alleles shared with modern Tuscans and modern Cornish are highlighted.

with the validation of ancient DNA sequences. What is necessary, however, is a set of standard criteria that everybody is reasonably happy about. Despite serious attempts to define these criteria (Cooper and Poinar 2000; Hofreiter et al. 2001), the present debate shows that a consensus has not been reached yet. As Helgason and Stefánsson (2003) remarked, errors can and do occur in large-scale DNA studies, either in the laboratory or in the construction of the databases. However, the impact of such errors, or of the possibility of such errors, should be evaluated critically before raising unjustified doubts about the conclusions of a study.

GUIDO BARBUJANI,¹ CRISTIANO VERNESI,¹
DAVID CARAMELLI,² LOREDANA CASTRÌ,³

CARLES LALUEZA-FOX,⁴ AND GIORGIO BERTORELLE¹
¹*Dipartimento di Biologia, Università di Ferrara, Ferrara, Italy;* ²*Dipartimento di Biologia Animale e Genetica, Laboratorio di Antropologia, Università di Firenze, Firenze, Italy;* ³*Dipartimento di Biologia Evoluzionistica e Sperimentale, Università di Bologna, Bologna, Italy;* and ⁴*Unitat de Biologia Evolutiva, Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona*

Electronic-Database Information

The URL for data presented herein is as follows:

Fluxus Engineering, <http://www.fluxus-engineering.com/sharenet.htm> (for Netwin 4.1)

References

- Bandelt H-J (2004) Etruscan artifacts. *Am J Hum Genet* 75: XXX–XXX (in this issue)
- Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743–753
- Bandelt HJ, Macaulay V, Richards M (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Mol Phylogenet Evol* 16:8–28
- Bandelt H-J, Quintana-Murci L, Salas A, Macaulay V (2002) The fingerprint of phantom mutations in mitochondrial DNA data. *Am J Hum Genet* 71:1150–1160
- Cooper A, Poinar HN (2000) Ancient DNA: do it right or not at all. *Science* 289:1139
- Handt O, Krings M, Ward RH, Pääbo S (1996) The retrieval of ancient human DNA sequences. *Am J Hum Genet* 59: 368–376
- Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, Stone A, Hedges R, Schaffner W, Utermann G, Sykes B, Pääbo S (1994) Molecular genetic analyses of the Tyrolean Ice Man. *Science* 264:1775–1778
- Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993) Toward a more accurate time scale for the human mitochondrial DNA tree. *J Mol Evol* 37:347–354
- Helgason A, Hickey E, Goodacre S, Bosnes V, Stefansson K, Ward R, Sykes B (2001) mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 68:723–737
- Helgason A, Stefansson K (2003) Erroneous claims about the impact of mitochondrial DNA sequence database errors. *Am J Hum Genet* 73:974–975
- Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. *Nat Rev Genet* 2:353–359
- Keyser-Tracqui C, Crubézy E, Ludes B (2003) Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol valley of Mongolia. *Am J Hum Genet* 73: 247–260
- Maca-Meyer N, Arnay M, Rando JC, Flores C, Gonzalez AM, Cabrera VM, Larruga JM (2004) Ancient mtDNA analysis and the origin of Guanches. *Eur J Hum Genet* 12:155–162
- Malyarchuk BA, Rogozin IB (2004) On the Etruscan mitochondrial DNA contribution to modern humans. *Am J Hum Genet* 75:XXX–XXX (in this issue)
- Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103–1110
- Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Menecier P, Hofreiter M, Possnert GG, Pääbo S (2004) No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol* 2:E57
- Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, Rengo C, Martinez-Cabrera V, Villems R, Kivisild T, Metspalu E, Parik J, Tolk HV, Tambets K, Forster P, Karger B, Francalacci P, Rudan P, Janicijevic B, Rickards O, Savontaus ML, Huoponen K, Laitinen V, Koivumaki S, Sykes B, Hickey E, Novelletto A, Moral P, Sellitto D, Coppa A, Al-Zaheri N, Santachiara-Benerecetti AS, Semino O, Scozzari R. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69:844–852
- Vernesi C, Caramelli D, Dupanloup I, Bertorelle G, Lari M, Cappellini E, Moggi-Cecchi J, Chiarelli B, Castri L, Casoli A, Mallegni F, Lalueza-Fox C, Barbujani G (2004) The Etruscans: a population-genetic study. *Am J Hum Genet* 74: 694–704
- Wakeley J (1993) Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. *J Mol Evol* 37:613–623

q10

Address for correspondence and reprints: Dr. Guido Barbujani, Dipartimento di Biologia, Università di Ferrara, via Borsari 46, I-44100 Ferrara, Italy. E-mail: bjg@dns.unife.it

© 2004 by The American Society of Human Genetics. All rights reserved.
0002-9297/2004/7505-00XX\$15.00

QUERIES TO THE AUTHOR

1 Au: Your article has been edited for grammar, clarity, consistency, and adherence to journal style. To expedite publication, we no longer ask authors for approval of routine grammatical and style changes. Please read the article to make sure your meaning has been retained; any layout problems (including table and figure placement) will be addressed after we have incorporated corrections. Note that we may be unable to make changes that conflict with journal style, obscure meaning, or create grammatical or other problems. If you are writing corrections by hand, please print clearly, and be aware that corrections written too close to the edges of the paper may not transmit by fax. Finally, please note that a delayed, incomplete, or illegible response may delay publication of your article. Thank you!

2 Au: OK to change “far” to “completely” here, as shown, or did you mean something different?

3 Au: I’ve deleted the phrase “in this way” from this sentence; is this change OK, or should it be replaced with something like “by making this comparison”?

4 Au: Correct to add a citation of Vernesi et al. (2004) here, as shown?

5 Au: Correct to add “a mutation at” before “16126” here, as shown?

6 Au: There is no Gilbert et al. (2003) in your reference list; could you please provide the complete reference information?

7 Au: Correct to add “is given” after “Bandelt et al. (1999),” as shown, or did you mean something different?

8 Au: Could you please verify the name of this software package, as well as the URL for the Web site (which I’ve added to an Electronic-Database Information section)? (The site itself seems to call the software “Network” rather than “Netwin.”)

9 Au: There is no Richards et al. (1998) or (1996) in your reference list; could you please provide the complete reference information?

10 Au: Did the editing improve the quality of your manuscript? Please select one: (1) Yes, very much; (2) Yes, somewhat; (3) No, not much; (4) No, not at all; (5) No opinion. What kind of change had the biggest impact on the quality of your manuscript? (L. Sarlo, Senior Manuscript Editor)

The American Journal of Human Genetics

The University of Chicago Press
1427 East 60th Street
Chicago, IL 60637

Please return this form even if no reprints are desired
 NO REPRINTS DESIRED

PAGE CHARGE / REPRINT ORDER FORM

(Please keep a copy of this document for your records)

AUTHORS: PLEASE RETURN REPRINT ORDER WITH CORRECTED PROOFS. LATE ORDERS WILL BE PRINTED AT ADDITIONAL COST. Please return this form to the above address immediately **even if no reprints are desired.** Authors are responsible for paying page charges for their articles. The attached form should be used to calculate page charges, color printing costs, and reprint costs. This form should be used to order reprints and pay page charges and color printing costs. Payment by check, money order, Visa, or MasterCard is required with all orders not accompanied by an institutional purchase order or purchase order number. Reprints ordered through an institution will not be processed without a purchase order number. Make checks and purchase orders payable to The University of Chicago Press. All purchase orders **MUST** include the following: journal name and date of issue, authors' names, title of article, and number of reprints ordered. Billing questions may be directed to Cindy Garrett, Billing Coordinator (tel.: [773] 753-8028; fax: [773] 753-0827; e-mail: Medical-Billing@press.uchicago.edu).

THE AMERICAN JOURNAL OF HUMAN GENETICS VOL. _____ NO. _____ MONTH _____

AUTHOR(S) NAME(S): _____

TITLE OF ARTICLE: _____

NUMBER OF PAGES: _____ MANUSCRIPT #: _____ COLOR IN ARTICLE? YES _____ NO _____

CHARGES: Please compute the total charges due.

Reprint charges

_____ # of Reprints ordered \$ _____

_____ # of Covers ordered \$ _____

SUBTOTAL \$ _____

Canadian GST \$ _____

(For orders shipped to Canada add 7% to above subtotal)

Shipping Fee \$ _____

(For non-U.S./non-Canadian orders add 45% to above subtotal)

Page charges \$ _____

(\$50.00 per page)

Color charges \$ _____

(\$650.00 for the first page, \$200.00 each additional)

TOTAL AMOUNT DUE (US \$) \$ _____

SHIPPING INFORMATION

Name _____

Phone _____ Fax _____

Dept _____ Room _____

Institution _____

Street _____

City _____ State _____ Zip _____

Country _____

BILLING INSTRUCTIONS (Institutional Orders Only)

Institution _____

Street _____

City _____ State _____ Zip _____

Country _____

PAYMENT DETAILS:

PLEASE NOTE: Orders will not be processed without a P.O. or credit card number or equivalent proof of payment.
MAKE CHECKS & PURCHASE ORDERS PAYABLE TO: The University of Chicago Press
Return this form with payment to the address at the top of this page.

All orders must be accompanied by one of three payment options (purchase order, check/money order, or Visa/MasterCard):

1) Purchase Order attached to come Institutional Purchase Order No. _____

Order will not be processed without a number

2) Check or Money Order for total charges is attached

3) Please charge to: VISA MasterCard

Card member name as it appears on card (please print clearly) _____

Card Number _____ Expiration date _____

Signature _____ Phone number _____

Signature: _____ **Date:** _____

Signature is required. By signing this form, the author agrees to accept responsibility for the payment of reprint and related costs described in this document.

The American Journal of Human Genetics

The University of Chicago Press
1427 East 60th Street
Chicago, IL 60637

PRICE LIST

AUTHORS: REPRINT ORDER MUST BE RECEIVED PRIOR TO PRINTING OF JOURNAL ISSUE. Authors are responsible for paying page charges for their articles. This form should be used to calculate page charges, color printing costs, and reprint costs. Payment by check, Money Order, Visa, or MasterCard is required with all orders not accompanied by an institutional purchase order or purchase order number. **Make checks and purchase orders payable to The University of Chicago Press.**

REPRINT CHARGES:

# OF PAGES	TOTAL QUANTITY				
	100	200	300	400	ADD'L 100
4	54.00	80.00	107.00	133.00	22.00
6	68.00	97.00	131.00	162.00	28.00
8	83.00	119.00	155.00	191.00	34.00
10	97.00	138.00	179.00	220.00	40.00
12	112.00	157.00	203.00	248.00	46.00
14	126.00	176.00	227.00	277.00	52.00
16	140.00	196.00	251.00	306.00	58.00
18	155.00	215.00	275.00	335.00	64.00
20	169.00	234.00	299.00	364.00	70.00
22	184.00	253.00	323.00	392.00	76.00
ADD'L 2 PAGES	13.00	18.00	23.00	28.00	5.00
COVERS	66.00	84.00	102.00	120.00	14.00

- For Canadian orders only add an addition 7% GST to the reprint cost. The 7% GST applies to all orders shipped to Canada regardless of the location of the purchasing institution or nationality of the recipient.
- For non-U.S. and non-Canadian orders only add 45% to the reprint cost for shipping charges.

PAGE CHARGES:

\$50.00 per page.

COLOR PRINTING CHARGES:

The first page costs \$650.00. Each additional page costs \$200.00.

Delivery

Orders will be shipped within 2-4 weeks of the journal print date. Allow extra time for delivery.