Letters to the Editor

Am. J. Hum. Genet. 70:1594–1596, 2002

Genetic Evidence for the Expansion of Arabian Tribes into the Southern Levant and North Africa

To the Editor:

In a recent publication, Bosch et al. (2001) reported on Y-chromosome variation in populations from northwestern (NW) Africa and the Iberian peninsula. They observed a high degree of genetic homogeneity among the NW African Y chromosomes of Moroccan Arabs, Moroccan Berbers, and Saharawis, leading the authors to hypothesize that "the Arabization and Islamization of NW Africa, starting during the 7th century AD, … [were] cultural phenomena without extensive genetic replacement" (p. 1023). H71 (Eu10) was found to be the second-most-frequent haplogroup in that area. Following the hypothesis of Semino et al. (2000), the authors suggested that this haplogroup had spread out from the Middle East with the Neolithic wave of advance. Our recent findings (Nebel et al. 2000, 2001), however, suggest that the majority of Eu10 chromosomes in NW Africa are due to recent gene flow caused by the migration of Arabian tribes in the first millennium of the Common Era (CE).

In the sample of NW Africans (Bosch et al. 2001), 16 (9.1%) of the 176 Y chromosomes studied were of Eu10 (H71 on a haplogroup 9 background). Of these 16 chromosomes, 14 formed a compact microsatellite network: 7 individuals shared a single haplotype, and the haplotypes of the other 7 were one or two mutational steps removed. This low diversity may be indicative of a recent founder effect. Where did these chromosomes come from?

The highest frequency of Eu10 (30%–62.5%) has been observed so far in various Moslem Arab populations in the Middle East (Semino et al. 2000; Nebel et al. 2001). The most frequent Eu10 microsatellite haplotype in NW Africans is identical to a modal haplotype (DYS19-14, DYS388-17, DYS390-23, DYS391-11, DYS392-11, DYS393-12) of Moslem Arabs who live in a small area in the north of Israel, the Galilee (Nebel et al. 2000). This haplotype, which is present in the Galilee at 18.5%, was termed the modal haplotype of the Galilee (MH Galilee) (Nebel et al. 2000). Notably, it is absent from two distinct non-Arab Middle Eastern populations, Jews and Muslim Kurds, both of whom have significant Eu10 frequencies—18% and 12%, respectively (Nebel et al. 2001). Interestingly, this modal haplotype is also the most frequent haplotype (11 [~41%] of 27 individuals) in the population from the town of Sena, in Yemen (Thomas et al. 2000). Its single-step neighbor is the most common haplotype of the Yemeni Hadramaut sample (5 [~10%] of 49 chromosomes; Thomas et al. 2000). The presence of this particular modal haplotype at a significant frequency in three separate geographic locales (NW Africa, the Southern Levant, and Yemen) makes independent genetic-drift events unlikely.

It should be noted that the Yemeni samples (Thomas et al. 2000) were not typed for the binary markers (p12f2 and M172) that define Eu10. However, both Yemeni modal haplotypes are present on a haplogroup background compatible with Eu10. These haplotypes carry a DYS388 allele with a high number of repeats (i.e., 17). High repeat numbers of DYS388, $\geqslant$15, were found to occur almost exclusively on Hg9, which comprises Eu9 and Eu10. Furthermore, in a sample of a six Middle Eastern populations, chromosomes with 17 repeats are frequent (40%) in Eu10 and rare (7%) in Eu9 (Nebel et al. 2001).

The term "Arab," as well as the presence of Arabs in the Syrian desert and the Fertile Crescent, is first seen in the Assyrian sources from the 9th century BCE (Eph'al 1984). Originally referring to nomads of central and northern Arabia, the term "Arabs" later came to include the sedentary population of the south, which had its own language and culture. The term thus covers two different stocks that became linguistically and culturally unified yet retained consciousness of their discrete origins (Grohmann et al. 1960; Rentz 1960; Caskel 1966, pp. 19–47; Goldziher 1967, pp. 45–97, 164–190; Beeston 1995; also see Peters 1999). Migrations of southern Arabian tribes northwards have been recorded mainly since the 3d century CE. These tribes settled in various places in central and northern Arabia, as well as in the Fertile Crescent, including areas that are now part of Israel (Dussaud 1955; Ricci 1984). The emergence of Islam in the 7th century CE furthered the unification of the Arabian tribal populations. This unified Arab-Islamic community engaged in a large movement of expansion, the Fertile Crescent and Egypt being the first areas to have

been conquered. It is very difficult to trace the tribal composition of the Muslim armies, but it is known that tribes of Yemeni origin formed the bulk of those Muslim contingents that conquered Egypt in the middle of the 7th century CE. Egypt was the primary base for raids further west into the Maghrib. The conquest of North Africa was difficult and took a few decades to complete (Abun-Nasr 1987). The region was militarily and administratively attached to Egypt until the beginning of the 8th century CE. Arab tribes of northern origin entered North Africa as well, both as troops and as migrants. A major wave of migration of such tribes, the Banu Hilal and Banu Sulaym, occurred during the 11th century CE (Abun-Nasr 1987). Thus, the Arabs, both southern (Yemeni) and northern, added to the heterogeneous Maghribi ethnic melting pot.

Little is known of the origins of the indigenous population of the Maghrib, the Berbers, except that they have always been a composite people. After the 8th century CE, a process of Arabization affected the bulk of the Berbers, while the Arab-Islamic culture and population absorbed local elements as well. Under the unifying framework of Islam, on the one hand, and as a result of the Arab settlement, on the other, a fusion took place that resulted in a new ethnocultural entity all over the Maghrib. In addition, Berber tribes sometimes claimed Arab descent in order to enhance their prestige. For example, the Berber nomadic tribe of the western Sahara, the Lamtuna, claimed descent from one of the South Arabian eponyms, Himyar. One of the chiefs of this Berber tribe, Lamtuna, is sometimes referred to as Saharawi, meaning "one of the nomads" or "one who comes from the Sahara" (Ibn al-Athir 1898, p. 462; Ibn Khallikan 1972, pp. 113, 128–129; Lewicki 1986). In Arabic sources, however, the name Saharawi is seldom used and does not seem to refer to a specific genealogical group. In light of these historical data, it is not surprising to find, among the Berbers and contemporary Saharawis of northern Africa, Y chromosomes that may have been introduced by recurrent waves of invaders from the Arabian Peninsula.

These documented historical events, together with the finding of a particular Eu10 haplotype in Yemenis, Palestinians, and NW Africans, are suggestive of a recent common origin of these chromosomes. Remarkably, the only non-Arabs in whom this haplotype has been observed to date are the Berbers (Bosch et al. 2001). It appears that the Eu10 chromosome pool in NW Africa is derived not only from early Neolithic dispersions but also from recent expansions from the Arabian peninsula.

Acknowledgments

We wish to thank Dr. Elena Bosch (University of Leicester, United Kingdom) for providing haplotype information on the NW African populations. This work was supported by a research grant from the Israeli Ministry of Science, Culture and Sport.

Almut Nebel,[1,*] Ella Landau-Tasseron,[2] Dvora Filon,[1] Ariella Oppenheim,[1] and Marina Faerman[3]

[1]*Department of Hematology, The Hebrew University–Hadassah Medical School and Hadassah University Hospital,* [2]*The Institute for Asian and African Studies, The Hebrew University, and* [3]*Laboratory of Biological Anthropology and Ancient DNA, The Hebrew University–Hadassah School of Dental Medicine, Jerusalem*

References

Abun-Nasr JM (1987) A history of the Maghrib in the Islamic Period. Cambridge University Press, Cambridge

Beeston AFL (1995) Saba. In: Bosworth CE, van Donzel E, Heinrichs WP, Lecompte G (eds) The encyclopedia of Islam, 2d ed. Vol 8. EJ Brill, Leiden, pp 663–664

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp continuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am J Hum Genet 68:1019–1029

Caskel W (1966) Gamharat an-nasab. Vol 1. EJ Brill, Leiden

Dussaud R (1955) La penetration des Arabes en Syrie avant l'-Islam. P Geuthner, Paris

Eph'al I (1984) The Ancient Arabs. The Magnes Press, The Hebrew University, Jerusalem

Goldziher I (1967) Muslim studies. Vol 1 (trans from German by Barber CR, Stern SM; ed by Stern SM). George Allen & Unwin, London

Grohmann A, Caskel W, Spuler B, Wiet G, Marcais G (1960) al-'Arab. In: Lewis B, Pellat C, Schacht J (eds) The encyclopedia of Islam, 2d ed. Vol 1. EJ Brill, Leiden, pp 524–533

Ibn al-Athir (1898) Annales du Maghreb et de l'Espagne. Alger

Ibn Khallikan (1972) Wafayat al-A'yan. Vol 7. Dar al-Thaqafa, Beirut

Lewicki T (1986) Lamtuna. In: Bosworth CE, van Donzel B, Lewis B, Pellat C (eds) The encyclopedia of Islam, 2d ed. Vol 5. EJ Brill, Leiden, p 652

Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. Am J Hum Genet 69:1095–1112

Nebel A, Filon D, Weiss D, Weale M, Faerman M, Oppenheim A, Thomas MG (2000) High-resolution Y chromosome haplotypes in Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. Hum Genet 107:630–641

Peters FE (ed) (1999) The Arabs and Arabia on the eve of Islam. Ashgate, Brookfield, VT

Rentz G (1960) Djazirat al-'Arab. In: Lewis B, Pellat C, Schacht J (eds) The encyclopedia of Islam, 2d ed. Vol 1. EJ Brill, Leiden, pp 533–555

Ricci L (1984) L'expansion de l'arabie meridionale. In: Chel-

hod J (ed) L'Arabie du Sud. Vol 1. Maisonneuve et Larose, Paris, pp 239–248

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens* in extant Europeans: a Y chromosome perspective. Science 290: 1155–1159

Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB (2000) Y chromosomes traveling south: the Cohen modal haplotype and the origins of the Lemba—the "Black Jews of South Africa." Am J Hum Genet 66:674–686

Address for correspondence and reprints: Dr. Marina Faerman, Laboratory of Biological Anthropology and Ancient DNA, Hebrew University–Hadassah School of Dental Medicine, Jerusalem 91120, Israel. E-mail: marinaf@pob.huji.ac.il

* Present affiliation: Division of Human Genetics, School of Pathology, National Health Laboratory Service (NHLS) and University of the Witwatersrand, Johannesburg, South Africa.

© 2002 by The American Society of Human Genetics. All rights reserved.
0002-9297/2002/7006-0023$15.00

Am. J. Hum. Genet. 70:1596–1598, 2002

SMN Dosage Analysis and Risk Assessment for Spinal Muscular Atrophy

To the Editor:

Feldkötter et al. (2002) recently reported a new method to determine, on the basis of real-time, quantitative PCR, copy numbers of *SMN1* (MIM 600354) and *SMN2* (MIM 601627). Their method allows a greater degree of automation and a faster turnaround time than do methods that have been described elsewhere (McAndrew et al. 1997; Chen et al. 1999; Wirth et al. 1999; Gérard et al. 2000; Scheffer et al. 2000; Ogino et al. 2001). Using their new method, they demonstrated that the copy number of *SMN2*—which is the centromeric homologue of *SMN1,* the disease gene for spinal muscular atrophy (SMA [MIM 253300 for type I; MIM 253550 for type II; and MIM 253400 for type III])—influences the severity of SMA in affected individuals with homozygous deletions of *SMN1*. They found that, the greater the copy number of *SMN2* was, the greater the likelihood was of a milder SMA type. Because this correlation is not absolute, they used Bayesian-type analyses to determine the posterior probabilities of developing each SMA type, with both a homozygous deletion of *SMN1* and a given copy number of *SMN2*. We discuss below several important ethical, prognostic, and technical issues raised in their article.

In table 6, Feldkötter et al. report "Probabilities That an Unaffected Who Has Been Tested after Birth and Has Been Found to Carry a Homozygous Absence of SMN1 Will Develop Type I, II, or III SMA, on the Basis of Number of SMN2 Copies." SMA is usually a childhood-onset disease, and testing of unaffected children is ethically problematic. We agree with the American Society of Human Genetics and the American College of Medical Genetics that "Timely medical benefit to the child should be the primary justification for genetic testing in children and adolescents" (American Society of Human Genetics Board of Directors and American College of Medical Genetics Board of Directors 1995, p. 1233). Since there are currently no effective treatments, presymptomatic or otherwise, for SMA, the timely medical benefit of the testing of unaffected children is unclear.

For the purpose of predicting SMA type from the *SMN2* copy number in unaffected children who lack *SMN1,* Feldkötter et al. perform Bayesian-type analyses by use of odds ratios, rather than conventional conditional probabilities. For the prior probabilities, they use the distribution of types of SMA among individuals affected with SMA: .51, for type I; .32, for type II; and .17, for type III. Even if one were to test unaffected children in this way, for this purpose, these prior probabilities would not be the correct ones to use for Bayesian or Bayesian-type analyses. If a child is asymptomatic at age 10 mo, for example, he or she is much less likely to have type I SMA than to have one of the other types (Zerres and Rudnik-Schöneborn 1995). One would have to incorporate the conditional probabilities of being asymptomatic at a particular age, for the hypothesis of each SMA type.

The data on *SMN2* copy number given by Feldkötter et al. could be used in prenatal testing, to predict SMA type. However, the prior probabilities that they use would be applicable only if the family history of SMA is of an unknown type. Although families with more than one type of SMA have been described—and are far from rare—knowing the type of SMA in an affected family member increases the prior probability of that type of SMA in a relative who is at risk of developing SMA. If the type of SMA in that affected family member is unknown, then the distribution of SMA types among all individuals with SMA would be relevant to the assignment of prior probabilities.

On the basis of all reported data, Feldkötter et al. state that, because two *SMN1* copies were found on 20/834 (2.4%) healthy chromosomes, "4.8% of normal individuals would be misinterpreted as noncarriers on the basis of the direct SMN1 test" (p. 365). Actually, these data imply that ~4.8% of noncarriers would have three copies of *SMN1* and that ~2.4% of carriers with an *SMN1* deletion on one chromosome 5 would have two *SMN1* copies on the other chromosome 5. We have referred to the latter as the "2 + 0" genotype (Chen et al. 1999). Taking into account the ~1.7% of carriers who

have an intragenic mutation undetectable as an *SMN1* exon 7 deletion, Feldkötter et al. state that this "reduces the sensitivity of the test to 93.5% for a person from the general population" (p. 365). Combining the ~1.7% of carriers who have an intragenic mutation with the ~2.4% (i.e., 0.024 × [1 − 0.017]) of carriers who have the 2 + 0 genotype gives the overall sensitivity of *SMN* dosage analysis for the detection of SMA carriers in the general population as ~95.9%. If an affected family member were known to have a homozygous deletion of *SMN1,* then the sensitivity of *SMN* dosage analysis for the detection of carriers among unaffected family members would be ~97.6% (i.e., 0.959/[1 − 0.017]). This is because the probability of an intragenic-mutation carrier in this family is greatly decreased relative to the probability of a 2 + 0 carrier (Ogino et al., in press).

Updating our combined data (McAndrew et al. 1997; Ogino et al. 2002, in press) gives 23 of 590 normal chromosomes 5 that have two copies of *SMN1*. Combining these data with those of Feldkötter et al. gives a total of 37 of 1,120 (3.3%) normal chromosomes 5 that have two copies of *SMN1*. We excluded other data in the literature (Wirth et al. 1999; Gérard et al. 2000; Scheffer et al. 2000), for reasons described elsewhere (Ogino et al., in press). On the basis of these numbers, ~3.2% (i.e., 0.033 × [1–0.017]) of carriers would have the 2 + 0 genotype. Therefore, the sensitivity of *SMN* dosage analysis for the detection of carriers in the general population would be ~95.1%, and that for the detection of carriers in a family with an affected individual lacking *SMN1* would be ~96.7% (i.e., 0.951/[1 − 0.017]).

Taking advantage of the single nucleotide differences between *SMN1* and *SMN2* in both exon 7 and intron 7, Feldkötter et al. used gene-specific primer pairs to amplify only *SMN1* or only *SMN2*. The primer pairs for each gene were mismatched for the other gene at either the final or the penultimate nucleotide from the 3′ end. These mismatches corresponded to the sequence differences in exon 7 (forward primers) and intron 7 (reverse primers). Gene conversions between *SMN1* and *SMN2,* which have been reviewed elsewhere (Burghes 1997), could potentially complicate this approach. If the *SMN1* exon 7 sequence (C) were converted to the *SMN2* exon 7 sequence (T) but the *SMN1* intron 7 sequence remained the same, the converted gene would presumably function as an *SMN2* gene in vivo. This is because the C→T transition in exon 7 of *SMN2,* although translationally silent, decreases the activity of an exonic splicing enhancer, so that less full-length protein is expressed (Lorson et al. 1999; Monani et al. 1999; Jong et al. 2000). By use of the gene-specific primers given by Feldkötter et al., the converted gene might have a different amplification efficiency from that of the normal *SMN1* or *SMN2* gene. Primers that are allele specific only for the functionally important polymorphism in exon 7 but not for the polymorphism in intron 7 might alleviate this problem.

SHUJI OGINO[1,2,3] AND ROBERT B. WILSON[4]
[1]*Department of Pathology, Brigham and Women's Hospital,* [2]*Department of Adult Oncology, Dana-Farber Cancer Institute, and* [3]*Harvard Medical School, Boston; and* [4]*Department of Pathology and Laboratory Medicine, University of Pennsylvania Medical Center, Philadelphia*

Electronic-Database Information

Accession numbers and the URL for data presented herein are as follows:

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for type I SMA [MIM 253300], type II SMA [MIM 253550], type III SMA [MIM 253400], *SMN1* [MIM 600354], and *SMN2* [MIM 601627])

References

American Society of Human Genetics Board of Directors, American College of Medical Genetics Board of Directors (1995) Points to consider: ethical, legal, and psychosocial implications of genetic testing in children and adolescents. Am J Hum Genet 57:1233–1241

Burghes AHM (1997) When is a deletion not a deletion? When it is converted. Am J Hum Genet 61:9–15

Chen KL, Wang YL, Rennert H, Joshi I, Mills JK, Leonard DG, Wilson RB (1999) Duplications and de novo deletions of the *SMNt* gene demonstrated by fluorescence-based carrier testing for spinal muscular atrophy. Am J Med Genet 85:463–469

Feldkötter M, Schwarzer V, Wirth R, Wienker TF, Wirth B (2002) Quantitative analyses of SMN1 and SMN2 based on real-time LightCycler PCR: fast and highly reliable carrier testing and prediction of severity of spinal muscular atrophy. Am J Hum Genet 70:358–368

Gérard B, Ginet N, Matthijs G, Evrard P, Baumann C, Da Silva F, Gerard-Blanluet M, Mayer M, Grandchamp B, Elion J (2000) Genotype determination at the survival motor neuron locus in a normal population and SMA carriers using competitive PCR and primer extension. Hum Mutat 16: 253–263

Jong YJ, Chang JG, Lin SP, Yang TY, Wang JC, Chang CP, Lee CC, Li H, Hsieh-Li HM, Tsai CH (2000) Analysis of the mRNA transcripts of the survival motor neuron (*SMN*) gene in the tissue of an SMA fetus and the peripheral blood mononuclear cells of normals, carriers and SMA patients. J Neurol Sci 173:147–153

Lorson CL, Hahnen E, Androphy EJ, Wirth B (1999) A single nucleotide in the *SMN* gene regulates splicing and is responsible for spinal muscular atrophy. Proc Natl Acad Sci USA 96:6307–6311

McAndrew PE, Parsons DW, Simard LR, Rochette C, Ray PN, Mendell JR, Prior TW, Burghes AH (1997) Identification of proximal spinal muscular atrophy carriers and patients by

analysis of SMN^T and SMN^C gene copy number. Am J Hum Genet 60:1411–1422
Monani UR, Lorson CL, Parsons DW, Prior TW, Androphy EJ, Burghes AH, McPherson JD (1999) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene *SMN1* from the copy gene *SMN2*. Hum Mol Genet 8: 1177–1183
Ogino S, Leonard DGB, Rennert H, Ewens WJ, Wilson RB. Genetic risk assessment in carrier testing for spinal muscular atrophy. Am J Med Genet (in press)
Ogino S, Leonard DGB, Rennert H, Gao S, Wilson RB (2001) Heteroduplex formation in *SMN* gene dosage analysis. J Mol Diagn 3:150–157
Ogino S, Leonard DGB, Rennert H, Wilson RB (2002) Spinal muscular atrophy genetic testing experience at an academic medical center. J Mol Diagn 4:53–58
Scheffer H, Cobben JM, Mensink RG, Stulp RP, van der Steege G, Buys CH (2000) SMA carrier testing—validation of hemizygous SMN exon 7 deletion test for the identification of proximal spinal muscular atrophy carriers and patients with a single allele deletion. Eur J Hum Genet 8:79–86
Wirth B, Herz M, Wetter A, Moskau S, Hahnen E, Rudnik-Schöneborn S, Wienker T, Zerres K (1999) Quantitative analysis of survival motor neuron copies: identification of subtle *SMN1* mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implications for genetic counseling. Am J Hum Genet 64:1340–1356
Zerres K, Rudnik-Schöneborn S (1995) Natural history in proximal spinal muscular atrophy. Clinical analysis of 445 patients and suggestions for a modification of existing classifications. Arch Neurol 52:518–523

Address for correspondence and reprints: Dr. Shuji Ogino, Department of Pathology, Brigham and Women's Hospital, 75 Francis Street, Amory 3d Floor, Boston, MA 02115. E-mail: sogino@partners.org
© 2002 by The American Society of Human Genetics. All rights reserved.
0002-9297/2002/7006-0024$15.00

Am. J. Hum. Genet. 70:1598–1599, 2002

Reply to Ogino and Wilson

To the Editor:

Drs. Ogino and Wilson (2002 [in this issue]) raised some issues regarding our paper on quantitative testing of *SMN1* and *SMN2* in spinal muscular atrophy (SMA) (Feldkötter et al. 2002). First, they raised some ethical issues regarding the testing of unaffected children for SMA. We are also aware of the controversial aspects of such testing and, in general, agree with Drs. Ogino and Wilson: the identification, at birth, of homozygous absence of *SMN1* in children, followed by the quantitative analysis of *SMN2,* should be offered as a prognostic tool only when a therapy for SMA is available. In this case, a newborn screening (similar to that in phenylketonuria) could—and possibly should—be considered. Since several drugs that up-regulate full-length *SMN2* have been found (Andreassi et al. 2001; Chang et al. 2001) and since the identification of many more is in progress, the development of a therapy for SMA seems likely to become a reality in the near future. Therefore, the development of a highly sensitive and fast method to determine the number of *SMN2* copies will be an essential prerequisite before starting a therapy. Furthermore, the identification, immediately after birth, of children who carry homozygous absence of *SMN1* will be equally essential, to start the therapy before the motor neurons are degenerated. On the basis of the number of *SMN2* copies, the dosage and starting-point of a therapy may significantly vary.

Since an efficient therapy has to be started early, we calculated the posterior probability that a child with an *SMN1* deletion would develop type I, type II, or type III SMA, under the assumption that the analysis is done immediately after birth. As a consequence, we have used a Bayesian-type analysis that is based on the odds ratios and a priori probabilities as chosen.

We reevaluated the sensitivity calculations, and we agree with Drs. Ogino and Wilson that the sensitivity of the test, for the detection of an SMA carrier from the general population without family history, is 95.9% (i.e., $1 - [0.024 + 0.017]$), since 2.4% of carriers have two *SMN1* copies per chromosome and 1.7% carry intragenic *SMN1* mutations. Therefore, there is a posterior probability of ~1:850 (i.e., [4.1/100] × [1/35]) that a person from the general population who carries two *SMN1* copies is an SMA carrier. The carrier frequency of 1:35 is based on the results presented in our previous article (Feldkötter et al. 2002). The sensitivity of the test for the detection of an SMA carrier from a family with an affected patient who carries a homozygous absence of *SMN1* is 97.6% (i.e., $1 - 0.024$).

With reference to the primers designed to detect either *SMN1* or *SMN2,* the test is based on two nucleotide differences in exon 7 and in intron 7 (position +100). This implies that converted *SMN* genes may amplify with a decreased efficiency. At this point, it is important to mention that, in the large majority (42/44 [~95%]) of converted *SMN* genes, the complete gene, except for the region containing the nucleotide difference in exon 8, is converted (Hahnen et al. 1996). This means that, for most converted *SMN* genes, the two primers that we have applied lie in either *SMN1* or *SMN2* only and will not hamper the efficiency of the PCR. Additionally, the analysis of 20 patients with only homozygous absence of *SMN1* exon 7 showed identical number of *SMN2* copies analyzed with both methods—multiplex competitive PCR (Wirth et al. 1999) and LightCycler PCR (Feldkötter et al. 2002). Nevertheless, the efficiency of the PCR may be reduced for those rarely observed *SMN*

genes in which the breakpoint lies between the two primers used in the LightCycler PCR.

MARKUS FELDKÖTTER,[1] VERENA SCHWARZER,[1] RADU WIRTH,[2] THOMAS F. WIENKER,[3] AND BRUNHILDE WIRTH[1]
[1]*Institute of Human Genetics,* [2]*Department of Surgery, and* [3]*Institute for Medical Biometry, Informatics and Epidemiology, University Clinic, Bonn*

References

Andreassi C, Jarecki J, Zhou J, Coovert DD, Monani UR, Chen X, Whitney M, Pollok B, Zhang M, Androphy E, Burghes AH (2001) Aclarubicin treatment restores SMN levels to cells derived from type I spinal muscular atrophy patients. Hum Mol Genet 10:2841–2849

Chang JG, Hsieh-Li HM, Jong YJ, Wang NM, Tsai CH, Li H (2001) Treatment of spinal muscular atrophy by sodium butyrate. Proc Natl Acad Sci USA 98:9808–9813

Feldkötter M, Wirth R, Schwarzer V, Wirth B (2002) Quantitative analysis of SMN1 and SMN2 based on real-time LightCycler PCR: fast and highly reliable heterozygosity testing and prediction of severity of spinal muscular atrophy. Am J Hum Genet 70:358–368

Hahnen E, Schönling J, Rudnik-Schöneborn S, Zerres K, Wirth B (1996) Hybrid survival motor neuron genes in patients with autosomal recessive spinal muscular atrophy: new insights into molecular mechanisms responsible for the disease. Am J Hum Genet 59:1057–1065

Ogino S, Wilson RB (2002) *SMN* dosage analysis and risk assessment for spinal muscular atrophy. Am J Hum Genet 70:1596–1598 (in this issue)

Wirth B, Herz M, Wetter A, Moskau S, Hahnen E, Rudnik-Schöneborn S, Wienker T, Zerres K (1999) Quantitative analysis of survival motor neuron copies: identification of subtle *SMN1* mutations in patients with spinal muscular atrophy, genotype-phenotype correlation, and implication for genetic counseling. Am J Hum Genet 64:1340–1356

Address for correspondence and reprints: Dr. Brunhilde Wirth, Institute of Human Genetics, University Clinic, Wilhelmstrasse 31, D-53111 Bonn, Germany. E-mail: bwirth@uni-bonn.de

© 2002 by The American Society of Human Genetics. All rights reserved.
0002-9297/2002/7006-0025$15.00

Am. J. Hum. Genet. 70:1599–1602, 2002

The Power of Multivariate Quantitative-Trait Loci Linkage Analysis Is Influenced by the Correlation between Variables

To the Editor:

In a recent article, Sham et al. (2000) investigated the power of variance-components linkage analysis by deriving an analytic expression for the noncentrality parameter (NCP) of the linkage test. The authors demonstrated that the NCP—and, hence, the power of the test to detect linkage—was determined primarily by the square of the additive and dominance genetic components of variance due to the quantitative-trait locus (QTL) and by the residual correlation between siblings. However, Sham et al. presented calculations for the univariate case only. Recently, it has been demonstrated that the power of QTL linkage analysis may be increased by use of multivariate techniques that analyze the pleiotropic action of the QTL on several variables (Boomsma 1996; Martin et al. 1997). In particular, the power of multivariate linkage analysis is strongly influenced by the correlation between the variables, being greatest when the QTL induces covariation between the variables in the direction opposite to the residual correlation (Allison et al. 1998; Amos et al. 2001). Here, I follow the methodology of Sham et al., to demonstrate analytically, for the first time, how the power of a bivariate variance-components linkage analysis depends not only on the magnitude and direction of the correlation between variables but also on the source of this correlation.

The relationship between two observable variables is parameterized in terms of the path model displayed in figure 1. Observed variables for each sib pair (*square boxes*) are due to the combined action of several latent variables (*circles*), including a pleiotropic QTL (Q), poly-

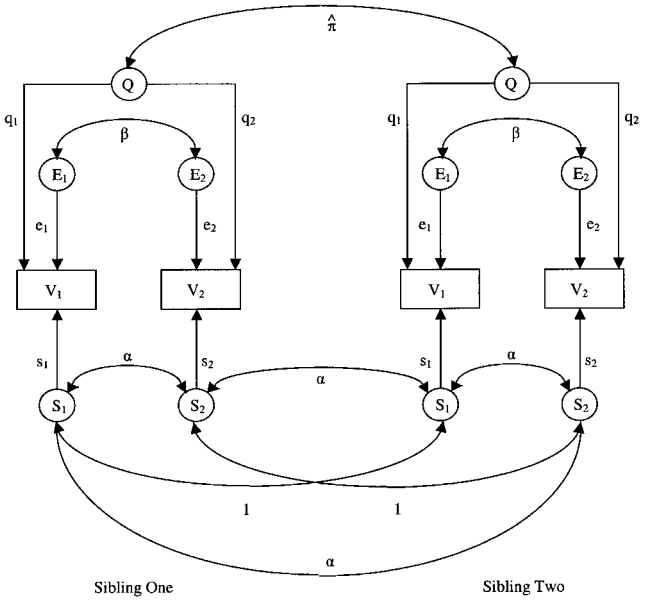

Figure 1 Path diagram showing the relationship between two observed variables (V_1 and V_2) for a pair of siblings. Covariation between the phenotypes is due to the QTL (Q), genetic and environmental sources that are shared among siblings (S_1 and S_2), and nonshared sources of variation (E_1 and E_2).

genic and environmental effects common to each member of the sib pair (S_1 and S_2), and unique environmental influences specific to each sibling (E_1 and E_2). Causal paths between variables are represented by unidirectional arrows, whereas correlations between variables are represented by bidirectional arrows. The strength of association between each variable is measured by a path coefficient (equivalent to a partial regression coefficient), in the case of a causal path, or a correlation coefficient, in the case of a bidirectional path. The correlation between siblings for the common QTL is $\hat{\pi}$, the estimated proportion of genes shared identical by descent at the trait locus, whereas the correlation between siblings for shared polygenic and environmental sources (i.e., S_1 and S_2) is 1. Correlations between phenotypes arise because of the pleiotropic action of the QTL (represented by the product of the path coefficients q_1 and q_2), from polygenic and environmental effects shared between siblings (represented by the product of α, s_1, and s_2) and from nonshared residual effects (represented by the product of β, e_1, and e_2). It is assumed that each variable is standardized to unit variance. The test for linkage is computed as twice the difference in log-likelihood between a model where q_1 and q_2 are estimated and a model where q_1 and q_2 are constrained to 0. Since q_1 (or, alternatively, q_2) is constrained to be positive, whereas q_2 has no such constraint (to allow for the possibility of a negative correlation between the observed variables), the test statistic is distributed asymptotically as a 50:50 mixture of χ^2_1 and χ^2_2 (Self and Liang 1987).

Under the null hypothesis of no linkage (N), the asymptotic parameter estimates for the covariance matrix, implied by figure 1, of the ith sib pair are

$$\Sigma_{i_N} = \begin{matrix} 1 & & & \\ q_1q_2 + \alpha s_1s_2 + \beta e_1e_2 & 1 & & \\ \frac{q_1^2}{2} + s_1^2 & \frac{q_1q_2}{2} + \alpha s_1s_2 & 1 & \\ \frac{q_1q_2}{2} + \alpha s_1s_2 & \frac{q_2^2}{2} + s_2^2 & q_1q_2 + \alpha s_1s_2 + \beta e_1e_2 & 1 \end{matrix}$$

(only lower elements of the matrix are shown). Under the alternative hypothesis of linkage (L), the asymptotic parameter estimates are given by:

$$\Sigma_{i_L} = \begin{matrix} 1 & & & \\ q_1q_2 + \alpha s_1s_2 + \beta e_1e_2 & 1 & & \\ \hat{\pi}_i q_1^2 + s_1^2 & \hat{\pi}_i q_1q_2 + \alpha s_1s_2 & 1 & \\ \hat{\pi}_i q_1q_2 + \alpha s_1s_2 & \hat{\pi}_i q_2^2 + s_2^2 & q_1q_2 + \alpha s_1s_2 + \beta e_1e_2 & 1 \end{matrix}$$

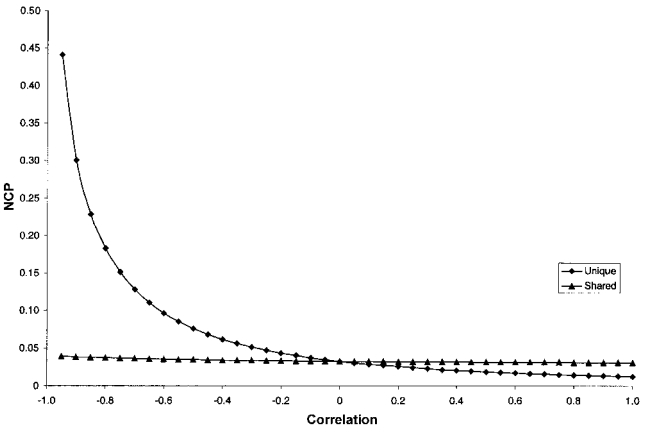

Figure 2 NCP as a function of either the correlation between unique sources of variation (*lines with diamonds*) or the correlation between shared sources of variation (*lines with triangles*).

According to Sham et al. (2000), the NCP for linkage (λ_L) is equal to twice the difference in expected log-likelihoods between the alternative and null hypotheses:

$$\begin{aligned}\lambda_L &= E(2\ln L_L) - E(2\ln L_N) \\ &= \ln|\Sigma_N| - \frac{1}{4}\ln|\Sigma_{\pi=0}| - \frac{1}{2}\ln|\Sigma_{\pi=0.5}| - \frac{1}{4}\ln|\Sigma_{\pi=1}| \ .\end{aligned}$$

To evaluate this expression, note that the determinant of a matrix of order n is a sum of $n!$ signed products, each involving n elements of the matrix. The computation is made easier, in the present case, because the variables are standardized and, therefore, the diagonal terms of the matrix are equal to 1:

$$\begin{aligned}|\Sigma| &= 1 + 2r_{21}r_{31}r_{32} + 2r_{21}r_{41}r_{42} + 2r_{31}r_{41}r_{43} \\ &\quad + 2r_{32}r_{42}r_{43} + r_{21}^2r_{43}^2 + r_{32}^2r_{41}^2 + r_{31}^2r_{42}^2 \\ &\quad - r_{21}^2 - r_{31}^2 - r_{32}^2 - r_{41}^2 - r_{42}^2 - r_{43}^2 \\ &\quad - 2r_{21}r_{32}r_{41}r_{43} - 2r_{21}r_{31}r_{42}r_{43} \\ &\quad - 2r_{31}r_{32}r_{41}r_{42} \ ,\end{aligned}$$

where r_{ij} is the element corresponding to the ith row and jth column of Σ. If we denote the right half of this equation as "$1 + x$" and note that the first-order Taylor-series expansion of $\ln(1 + x) \approx x$, then the NCP may be approximated as

$$\lambda_L \approx x_S - \frac{1}{4}x_2 - \frac{1}{2}x_1 - \frac{1}{4}x_0 \ ,$$

where x_S, x_2, x_1, and x_0 are the first-order Taylor-series approximations for the null hypothesis and the alternative

hypotheses of sharing two, one, or zero alleles identical by descent at the trait locus.

Evaluation of this expression in terms of the parameters in figure 1 yields

$$\begin{aligned}\lambda_L \approx{}& \frac{q_1^4}{8}+\frac{q_2^4}{8}+\frac{q_1^2q_2^2}{4}+\frac{q_1^4q_2^4}{2}\\ &-\frac{q_1^2q_2^4}{2}-\frac{q_1^4q_2^2}{2}-\frac{q_1^4s_2^4}{8}\\ &-\frac{q_2^4s_1^4}{8}-\frac{q_1^2q_2^2s_1^2s_2^2}{4}\\ &+\frac{\alpha q_1q_2^3s_1s_2(q_1^2+s_1^2-1)}{2}\\ &+\frac{\alpha q_1^3q_2s_1s_2(q_2^2+s_2^2-1)}{2}\\ &+\frac{\beta q_1q_2^3e_1e_2(q_1^2-1)}{2}+\frac{\beta q_1^3q_2e_1e_2(q_2^2-1)}{2}\\ &+\alpha\beta q_1^2q_2^2s_1s_2e_1e_2+\frac{\beta^2q_1^2q_2^2e_1^2e_2^2}{2}\ .\end{aligned}$$

Note particularly that the second part of the equation (i.e., the last four lines) contains terms involving the correlation between shared polygenic and environmental effects (α) and the correlation between unique environmental effects (β). The sign of these correlations contributes to the magnitude of the NCP. Consider first the terms containing the correlation between shared polygenic and environmental effects (i.e. the terms containing α). It is apparent that the parts of the expression inside parentheses must be negative. Therefore, if the QTL and shared polygenic and environmental influences produce correlations in the same direction, the terms will be negative, and therefore the NCP and the power to detect linkage will decrease. In contrast, when the QTL and shared influences induce correlations in opposite directions, the terms will become positive increasing the NCP and power. The power to detect linkage increases as the correlation between shared sources decreases (i.e., becomes more negative). A similar argument also applies to terms containing the QTL and unique sources of variation (i.e., the terms that include α), although the increase in power is more dramatic because the terms inside the parentheses are greater in magnitude and because there is an additional term containing β that is always positive (i.e., $1/2\beta^2q_1^2q_2^2e_1^2e_2^2$). The last term in the equation (i.e., $\alpha\beta q_1^2q_2^2s_1s_2e_1e_2$) suggests that the increase in power will be greatest when both shared and nonshared influences induce covariance in the direction opposite to the QTL.

Figure 2 displays the effect that varying the correlation between shared and unique sources of variation has on the NCP for a plausible biological model. In this model, the QTL accounts for 20% of the variance of each trait (i.e., $q_1^2 = q_2^2 = 0.2$), and induces a positive correlation between the variables (i.e., q_1 and q_2 are both positive). Both shared and unique effects account for forty percent of the variance for both traits (i.e., $s_1^2 = s_2^2 = 0.4$; $e_1^2 = e_2^2 = 0.4$). The correlation between unique sources of variation is varied, while the shared correlation is fixed at 0 (*lines with diamonds*), and the correlation between shared factors is varied, whereas the unique environmental correlation is fixed at 0 (*lines with triangles*). Note that the graph is based on exact values for the NCP and not on the Taylor-series approximation.

In both cases, the NCP increases as the correlation between the latent sources of variation decreases. However, although the increase in NCP is small and linear for the shared case, the increase is dramatic and exponential as the correlation between the unique sources of variation decreases. Thus, the power of bivariate QTL linkage analysis depends not only on the phenotypic correlation between variables but also on the source of this correlation.

In conclusion, these results imply that, in a bivariate linkage analysis, one is most likely to detect a QTL that produces a correlation between variables opposite in direction to the background correlation. In particular, power is dramatically affected by the correlation between the unique environmental sources of variation. This combination of latent sources would tend to produce variables that have low or moderate phenotypic correlations, a fact that should be kept in mind when deciding which variables to include in a bivariate linkage analysis.

Acknowledgments

I would like to thank Dr. David Duffy for fruitful discussions and Dr. Nick Martin for helpful comments on the manuscript.

DAVID M. EVANS
Queensland Institute of Medical Research and Joint Genetics Program
University of Queensland
Brisbane

References

Allison DB, Thiel B, St Jean P, Elston RC, Infante MC, Schork NJ (1998) Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. Am J Hum Genet 63:1190–201

Amos C, de Andrade M, Zhu D (2001) Comparison of multivariate tests for genetic linkage. Hum Hered 51:133–144

Boomsma DI (1996) Using multivariate genetic modeling to detect pleiotropic quantitative trait loci. Behav Genet 26: 161–166

Martin N, Boomsma D, Machin G (1997) A twin-pronged attack on complex traits. Nat Genet 17:387–392

Self SG, Liang K (1987) Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J Am Stat Assoc 82:605–610

Sham PC, Cherny SS, Purcell S, Hewitt JK (2000) Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. Am J Hum Genet 66:1616–1630

Address for correspondence and reprints: Dr. David M. Evans, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane 4029, Australia. E-mail: davidE@qimr.edu.au

© 2002 by The American Society of Human Genetics. All rights reserved.
0002-9297/2002/7006-0026$15.00

Am. J. Hum. Genet. 70:1602, 2002

The National Institutes of Health Announces Online Availability of "Points to Consider When Planning a Genetic Study That Involves Members of Named Populations"

To the Editor:

The National Institutes of Health (NIH) has developed a guide for researchers, called "Points to Consider When Planning a Genetic Study That Involves Members of Named Populations." The NIH supports and encourages the concept and process of community consultation in many research areas and believes that investigators who are planning genetic-research projects involving members of named populations should consider whether and how those communities should be consulted. The new "Points to Consider" document describes what is meant by "community consultation"; presents situations in which community consultation should be considered; identifies potential benefits, both for researchers and for communities, that engagement in this process offers; and provides practical examples of how to plan a community consultation. The "Points to Consider" document is posted on the NIH Web site, at the URL given below.

It is increasingly important for researchers to realize that nonscientists may not be well versed in the scientific benefits resulting from genetics research. Individuals and the communities to which they belong may fear that participation in genetic studies involving named populations may end up stereotyping that particular named population, potentially putting the entire community at risk of discrimination by insurers or other third parties. In creating the "Points to Consider" document, the NIH aims to assist scientists in the design of studies that operate in variable social and cultural contexts and that yield meaningful data while they work with communities.

JUDITH H. GREENBERG
National Institute of General Medical Sciences
National Institutes of Health
Bethesda, MD

Electronic-Database Information

The URL for data in this letter is as follows:

Points to Consider When Planning a Genetic Study That Involves Members of Named Populations, http://www.nih.gov/sigs/bioethics/named_populations.html

Address for correspondence and reprints: Dr. Judith H. Greenberg, National Institutes of Health, Division of Genetics and Developmental Biology, Building 45, Room 2As25, 45 Center Drive, MSC-6200, Bethesda, MD 20892-6200. E-mail: greenbej@nigms.nih.gov

This article is in the public domain, and no copyright is claimed.
0002-9297/2002/7006-0027