

In a Vietnamese population, *MSX1* variants contribute to cleft lip and palate

Yasushi Suzuki,* DDS¹, Peter A. Jezewski,* DDS, MA², Junichiro Machida, DDS, PhD^{1,3}, Yoriko Watanabe, MD, PhD⁴, Min Shi, MS⁴, Margaret E. Cooper, MS, MSIS, MEd⁵, Le Thi Viet, DDS⁶, Nguyen Thi Duc Tin, MD⁶, Huynh Hai, DDS⁶, Nagato Natsume, DDS, PhD¹, Kazuo Shimozato, DDS, PhD¹, Mary L. Marazita, PhD^{5,7}, and Jeffrey C. Murray, MD^{4,8}

Purpose: To identify causes of nonsyndromic cleft lip and palate in a Vietnamese population. **Methods:** In this study, 175 families with at least one case of cleft lip and/or palate were studied using the candidate genes *TGFA*, *MSX1*, and *TGFB3*. **Results:** Transmission distortion for alleles of *MSX1* were demonstrated for the whole population and two missense mutations were identified, including one (P147Q) that is found in approximately 2% of the population. The P147Q appears to arise from a founder individual based on shared haplotypes in unrelated families. **Conclusions:** *MSX1* contributes to nonsyndromic clefting in a Vietnamese population, and consistent with other studies, identifiable mutations in this gene cause about 2% of cases of nonsyndromic clefting. **Genet Med** 2004;6(3):117–125.

Key Words: cleft lip, cleft palate, *MSX1*, *TGFA*, *TGFB3*

Cleft lip and/or palate is one of several complex human birth defect traits that are thought to be caused by a combination of genes and environmental interactions.¹ Segregation analyses and twin studies support a genetic role for cleft lip and palate and a large number of common as well as uncommon teratogens, including cigarette smoke and alcohol, have been suggested to contribute to the environmental component.² Clefting is a ubiquitous disorder world wide, although its incidence varies widely from highest in Asian populations to lowest in those of African descent.³ A variety of genetic and epidemiologic approaches have been used to identify etiologies with candidate gene studies having provided the strongest suggestive results to date. Selection of candidate genes is often based on their phenotypes from mouse knockouts or their expression patterns in human or animal embryos. A role for transforming growth factor α (*TGFA*), transforming growth factor

β_3 (*TGFB3*), and the muscle segment homeobox gene (*MSX1*) has been supported by one or more association studies.^{4–7} Among these candidate genes, it is the *MSX1* association that has most often been replicated. A family identified in which a stop mutation in the first exon of *MSX1* cosegregates in individuals with cleft lip and palate or isolated cleft palate first confirmed *MSX1* as having a direct role in human clefting disorders.⁸ We subsequently demonstrated mutations in about 2% of nonsyndromic cleft cases in a panethnic collection with a range of missense mutations and variants in conserved elements.⁹ In the current study, we used a combination of the family-based association test method, a log-linear method for triads stratified on parental mating type and direct sequencing to evaluate the role of three major candidate genes in individuals of Vietnamese descent.

MATERIALS AND METHODS

Subjects

The subjects of this study were 168 sporadic cases of nonsyndromic cleft lip with or without cleft palate, and isolated cleft palate. Vietnamese patients collected as cases with their parents including 49 CLO (cleft lip only), 106 CLP (cleft lip and/or palate), and 13 CPO (cleft palate only) patients. In addition, seven cases were collected (1 CLO, 3 CLP, and 3 CPO) with a positive family history bringing the total families in this study to 175. These groups are further characterized by laterality and gender as shown in Table 1. No further clinical data are available for these families beyond that reported here.

Blood spots on filter cards were collected from 1996 to 2002 through the Japanese Cleft Palate Foundation, a medical organization that provides clinical care for underserved popula-

¹Second Department of Oral and Maxillofacial Surgery, School of Dentistry, Aichi-Gakuin University, Nagoya, Japan; ²The Forsyth Institute, Department of Cytokine Biology and Harvard School of Dental Medicine, Department of Oral and Developmental Biology, Boston Massachusetts; ³Division of Oral and Maxillofacial Surgery, Toyohashi Municipal Hospital, Toyohashi, Japan; ⁴Department of Pediatrics, Division of Neonatology, University of Iowa, Iowa City, Iowa; ⁵Center for Craniofacial and Dental Genetics, The University of Pittsburgh, School of Dental Medicine, Pittsburgh, Pennsylvania; ⁶Odonto-Maxillo-Facial Center, Ho Chi Minh City, Vietnam; ⁷Department of Oral and Maxillofacial Surgery, School of Dental Medicine; Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pennsylvania.

*Both authors contributed equally to this study.

Jeffrey C. Murray, MD, University of Iowa, Department of Pediatrics, Division of Neonatology, 2182 ML, Iowa City, IA 52242.

Received: August 14, 2004.

Accepted: January 30, 2004.

DOI: 10.1097/01.GIM.0000127275.52925.05

Table 1

Characteristics of the nonsyndromic orofacial cleft case families

	Total	CLO	CL+/-P	CPO
No. of triads (proband)	175 (176) ^a	50 (51) ^a	109 (109)	16 (16)
Sporadic	168	49	106	13
Familial	7	1	3	3
Laterality				
Right side cleft		14	27	
Left side cleft		27	47	
Bilateral cleft		9	30	
Unknown		1	5	
Gender				
Male		23	62	6
Female		27	47	10
Unknown		1		

^aThere was one set of twin brothers, both with CLO among the set of case triads. One brother had bilateral CLO, whereas the other had incomplete L side CLO.

tions in Japan, Vietnam, Mongolia,¹⁰ as well as through the local hospitals: Odonto-Maxillo-Facial-Center of Hochiminh City in Ho Chi Minh City, Nguyen dinh Chieu Hospital in Bentre province, and Ninh Binh General Hospital in Ninh Binh province. DNA was extracted from blood spots using the DNeasy 96 Tissue Kit (Qiagen, Chatsworth, CA). This project obtained Aichi-Gakuin University-approved informed consent from all subjects.

PCR and genotyping

Approximately 40 ng of template genomic DNA was analyzed by polymerase chain reaction (PCR). PCR was performed in 1× PCR buffer [10 mmol/L Tris-HCl (pH8.3), 1.5 mmol/L MgCl₂, 50 mmol/L KCl] with 200 μmol/L dNTPs, and 0.12–0.2 μmol/L primers (final concentration). Thermocycling was performed with an initial 3 minute denaturation at 94°C followed by 35 cycles of: 30 seconds at 94°C, 30 seconds at annealing temperature and 30 seconds at 72°C, and final extension at 72°C for 3 minutes. To detect CA repeat polymorphisms in *MSX1* and *TGFB3*, single strand conformational analysis (SSCP) was performed. The amplified products were separated on MDE (Mutation Detection Enhancement from BioWhittaker Molecular Applications) nondenaturing gels for 5–6 hours at 20 W at room temperature with fans. DNA bands were visualized by silver staining and inspected for potential variants. The *MSX1* CA repeat is found within the single intron of the *MSX1* gene. Using the AF426432 Genbank entry as a reference, it is found at position 2450 and is shown therein as nine CAs. The *MSX1* CA4 allele is represented by the PCR product with nine repeats and is the most common allele in all populations studied to date. The primers used for PCR amplification and sequencing of genomic DNA are given in Tables 2 and 3.

Kinetic PCR reactions for TGFA C3827T¹¹ genotyping were performed on GeneAmp 7900 Sequence Detection System (PE Applied Biosystems) in duplicates with each of the two allele specific primers in a total volume of 3 μL for each reaction. 1× SYBR Green Master Mix (Applied Biosystems No. 4309155) was used in PCR reaction with a final concentration of 0.2

Table 2

Sequences of primers used for PCR

Gene	Region	Primer position	Primer sequence for PCR 5' to 3'	Product size	Tm (C)
<i>MSX1</i>	exon 1	414–433 ^a	M13F ^b –TGGCCAGTGCTGCGGCAGAA	841 bp	60
		1246–1264	M13R ^c –ATTCATCCGCTGGGGTGAA		
	intron 1	1490–1508	GGGCACCAAGTCAATTAGG	661 bp	58
		2129–2148	TCCCGGCAGAAAGTGAATTC		
	CA repeat	2385–2404	TCTGCTGACTATGCCACAGG	190 bp	55
		2554–2573	CGAGGGGCTGCTTTATTCTA		
<i>TGFB3</i>	CA repeat	3133–3152	ACTTGCGGCACTCAATATC	701 bp	55
		3814–3833	CTCAGGGAGCAAAGAGGTGA		
		–61129 ^d to –61148	AGATTCTGGCTTCCACGAAA	216 bp	55
TGFA	T3827C	–61323 ^d to –61344	GCAAGCAGGGATAATAACAGCA		
		576–604 ^e	GAGAAGCTGTATCCTCTAACCACGAGAAT	78 bp	56
		577–603	AGAAGCTGTATCCTCTAACCACGAGAAC		
		634–653	GGATGGTCTTCAATGTCATG		

^aPosition of the primers for the *MSX1* PCR reactions are nucleotide positions within the genbank entry AF426432.

^bM13F sequence is TGTAACACGACGGCCAGT.

^cM13R sequence is CAGGAACAGCTATGACC.

^dPosition of the *TGFB3* CA primers relative to the ATG start codon of *TGFB3*.

^ePosition of the TGFA kinetic primers are given within the genbank entry AF123243.

Table 3
Sequences of primers used in *MSX1* sequencing analysis

Region	Primer position ^a	Primer sequence for sequence (5' to 3')
exon 1	414–433	TGGCCAGTGCTGCGGCAGAA
	460–479	GGCTGCTGACATGACTTCTT
	921–940	CCTGGGTCTGGCTACTCAC
	M13(–21) ^f	TGTAAACGACGCGCCAGT
	648–666	TCATGGCCGACCACAGGAAC
intron 1	921–940	CCTGGGTCTGGCTACTCAC
	1490–1508	GGGCACCAAGTCAATTAGG
exon 2	2090–2109	ACGCGGGTTTTCGCTAAGTA
	3225–3245	TTTTTTTCTTTTCGCCCTCA
	3722–3739	ACCAGGGCTGGAGGAATC
	3510–3529	AAGGCAGCCGGTGGCAGCAT
	3814–3833	CTCAGGGAGCAAGAGGTGA

^aPosition is nucleotide number within Genbank entry AF426432.

μmol/L for each primer and 0.25 μL of blood spot DNA in each reaction. PCR was performed with an initial enzyme activation at 95°C for 10 minutes, followed by 55 cycles of 20 seconds at 95°C for denaturation, and 20 seconds at 56°C for annealing and extension. Primers for the *TGFA* kinetic PCR reaction are given at our web site <http://genetics.uiowa.edu/publications/rschultz/primerseq.html>.

Haplotype analyses for *MSX1* were performed with TaqMan assays for common single nucleotide polymorphisms (snps). SNPs 2–5 from Table 4 were discovered as a result of a dense search for disease-associated snps at the *MSX1* locus. These were converted to TaqMan Assays-by-Design by Applied Biosystems. The collection of 48 rare and common snps found at the *MSX1* locus from previous work⁹ was published online, January 2002, at <http://genetics.uiowa.edu/publications/peterj/ElectronicAppendix2B.html> and in Table 3. These snps and other snps found independently,¹² can serve as a resource for the research community. Snp1 and 2 were found online (http://www.appliedbiosystems.com/products/productdetail.cfm?prod_id=1141) as TaqMan Assays-on-Demand. The

Table 4
Single nucleotide polymorphism TaqMan assays

SNP no.	Type/ID ^a	Position
snp1	AOD/C_2402343_10	–8796A > G ^b
snp2	AOD/C_2876136_10	–5066A > G ^b
snp3	AbD	451 + 1201T > G, 2121 ^c
snp4	AbD	452 – 667T > G, 2580 ^c
snp5	AbD	452 – 402G > T, 2845 ^c

^aAOD is assay on demand; AbD is assay by design; both from Applied Biosystems.

^bPosition is relative to the ATG start codon of *MSX1*.

^cPosition is within Genbank entry AF426432.

TaqMan genotyping was performed on a GeneAmp 7900 Sequence Detection System in 3 μL reaction volumes, in duplicate. See Table 4 for TaqMan single nucleotide polymorphism details. PCR reactions were performed in 1× TaqMan Universal PCR Master Mix (part no. 4304437), from Applied Biosystems, 1× of the Assay-On-Demand/Assay-By-Design primer assay mix, and 2 ng of genomic DNA. PCR was performed with an initial enzyme activation at 95°C for 10 minutes followed by 40 cycles of 92°C for 15 seconds and 60°C for 1 minute.

Direct sequencing

Sequencing was performed on the *MSX1* gene only after tests of association supported mutation searches in this gene. PCR products were used for sequencing reactions directly or after purification by extraction from 2% agarose gels using Qiagen Gel Purification kit (Qiagen, Chatsworth, CA). Cycle sequencing was performed according to the manufacturer's instructions using ABI PRISM BigDye Terminator Sequencing kits with AmpliTaq DNA polymerase FS (Perkin Elmer, Foster City, CA) on an Applied Biosystems DNA Sequencing System (model 373 or ABI3700 sequencer). Data were collected and analyzed with the Phred/Phrap/Consed System to identify sequence variants (<http://www.phrap.org>).

Data analysis

For the transmission disequilibrium tests (TDT), the Family-Based Association Test¹³ was used. The association model was additive with the degrees of freedom = no. of alleles – 1. The biallelic mode compares each allele to all others, whereas the multiallelic mode simultaneously tests the distribution of all alleles. *P* values for biallelic TDT were not corrected for multiple comparisons.

The relative risk estimate for the *MSX1* CA4 allele was analyzed using a likelihood ratio test (LRT).¹⁴ Log-linear models are used to estimate the relative increase in risk associated with having one or two copies of the “high-risk” allele. The *MSX1* CA data were divided into the normal (1 = CA 1, 2, 3, 5) and variant alleles (2 = CA 4) because this division was shown to be significantly relevant from the FBAT analysis. This data were analyzed with SAS, V8 using the GENMOD procedure.

Haplotype reconstruction from the Vietnamese control population was performed with PHASE, version 1.0 (<http://archimedes.well.ox.ac.uk/pise/PHASE-simple.html>), using the Advanced PHASE form.

RESULTS

Candidate gene evaluation by Family-Based Association Tests

TGFA

TGFA was analyzed using a kinetic PCR assay for a common point polymorphism located in the 3' untranslated region,¹¹ which has been shown to be in linkage disequilibrium (LD) with the less frequent Taq1 polymorphism (A. Vieira and J. Murray, personal communication, 2004), a polymorphism that had been initially reported as being associated with cleft lip

Table 5
Full variant list for *MSX1* Vietnamese samples

Nucleotide position ^a	Original	Nucleotide variant	Amino acid variant	93 Probands (2001); 100 samples	124 Probands (2002); 126 samples	2002 & 2001 totals
				Total no.	Total no.	Total no.
5'UTR						
−30G > A	G/G	G/A		1	0	1
Exon 1						
330C > T	C/C	C/T	G110G	15	8	23
330C > T		T/T	G110G	2	3	5
159G > C	G/G	G/C	L53L	1	0	1
101C > G	C/C	C/G	A34G	4	9	13
293G > A	G/G	G/A	G98E	2	2	4
440C > A	C/C	C/A	P147Q	5	2	7
Intron						
451 + 3G > A	G/G	G/A		2	0	2
451 + 34G > A	G/G	G/A		0	1	1
451 + 232A > T	A/A	A/T		7	22	29
	A/A	T/T		0	29	29
451 + 219G > A	G/G	G/A		0	8	8
		A/A		0	1	1
451 + 590G > C	G/G	G/C		35	25	60
		C/C		21	25	46
Exon 2						
nothing				0	0	0
3'UTR						
*6C > T	C/C	C/T		2	8	10
	C/C	T/T		0	1	1

^aNomenclature for the position of the variants is by reference to the start and stop codons as well as the exon boundaries.²⁶

and palate.¹⁵ In this study, we performed transmission distortion analysis and the data are shown in Tables 6 and 7. No evidence of transmission distortion could be identified with *TGFA* in this population.

TGFB3

TGFB3 was studied using a CA repeat allele that has been previously evaluated in Filipino and related populations.¹⁶ No evidence of transmission distortion was observed in the total population of triads ($n = 175$) for any of the alleles of *TGFB3* studied. However, there was borderline significance among the sporadic families ($n = 168$) for the 1 and the 4 allele (Table 6).

MSX1

MSX1 was evaluated using the CA repeat initially defined in Padanilam et al.,¹⁷ which has shown evidence of LD with cleft lip and palate previously in populations drawn from Iowa, Maryland, and South America.^{5,18,19} Transmission distortion testing performed on only completely typed triads of mother/

father/affected child showed a significant transmission distortion in a biallelic model, with overrepresentation of the 4 allele ($P = 0.022$) as shown in Table 6. Given that 94% of all the *MSX1* CA alleles were either the CA2 or the CA4, this is essentially a two allele system. Thus, as expected the CA2 allele was shown to be significantly undertransmitted in a biallelic model ($P = 0.021$). The multiallelic model did not show significant transmission distortion (Table 7; $P = 0.09$), because the significant positive association with allele 4¹⁶⁹ was balanced by the significant negative associations with allele 2¹⁷³, leading to nonsignificant multiallelic P values. Note that for *MSX1*, as well as for *TGFB3*, the sporadic group of families ($n = 168$) had slightly more significant results (i.e., lower P values) than those results obtained with the total set of triads including those with familial history ($n = 175$).

Evaluation by LRT and sequencing

To further evaluate the potential role of the *MSX1* CA4 allele in the disease association, log-linear modeling was performed.

Table 6
Transmission Disequilibrium Test results (mode: biallelic)

Marker	Allele name ^{a,b}	All families (<i>n</i> = 175)			Sporadic families (<i>n</i> = 168)		
		T	NT	<i>P</i> value	T	NT	<i>P</i> value
<i>MSX1</i>	1/175/12 ^a	6	5	0.763	6	5	0.763
<i>MSX1</i>	2/173/11	46	71	0.021 (–)	44	71	0.012 (–)
<i>MSX1</i>	3/171/10	0	2	0.157	0	2	0.157
<i>MSX1</i>	4/169/9	72	47	0.022 (+)	72	45	0.013 (+)
<i>MSX1</i>	5/167/8	2	1	0.564	2	1	0.564
TGFB3	1/166C ^b	85	104	0.167	77	103	0.053 (–)
TGFB3	3/170C	1	0	0.317	1	0	0.317
TGFB3	4/168C	104	87	0.219	103	79	0.075 (+)
TGFB3	5/172C	1	0	0.317	1	0	0.317
TGFA	C	68	57	0.325	66	56	0.325
TGFA	T	57	68	0.325	56	66	0.325

T, number of transmitted alleles.

NT, number of nontransmitted alleles; *P* values less than or equal to 0.10 are bolded. If *P* values are borderline, a + or – sign is listed, where + means there is a positive association with the allele, the allele is transmitted more than expected, and – means the allele is transmitted less often than expected.

^aThe names of the *MSX1* CA repeat alleles are in the format of: allele number/length of PCR product in base pairs/number of CA repeats.

^bNames of the alleles for TGFB3 are derived as follows: 1=9CAs+C (166C), 2=10CAs+T (168T), 3=11CAs+C (170C), 4=10CAs+C (168C), 5=12CAs+C (172C).

Table 7
Family-Based Association Test results (mode: multiallelic)

Marker		No. of alleles	DF	CHISQ	<i>P</i>
<i>MSX1</i>	ALL	5	4	8.018	0.091
	Sporadic	5	4	9.046	0.06
TGFB3	ALL	4	3	3.910	0.271
	Sporadic	4	3	5.75	0.124
TGFA	ALL	2	1	0.968	0.325
	Sporadic	2	1	0.870	0.365

Test parameters used were association model as additive with df = no. of alleles – 1; test mode is multiallelic.

Of the 168 sporadic triads, 162 had full triad *MSX1*-CA genotype data available. These were stratified according to the mating type, which is defined by the number of variant alleles carried by each of the two parents. These data are shown in Table 8 and were analyzed with SAS under three sets of assumptions as shown in Table 9. Under the assumption of Hardy-Weinberg Equilibrium, (HWE), the analyses of the data indicate significance for both a single and a double dose of the CA4 allele. This same pattern is nearly significant under no assumption of HWE. It also appears that the analyses of the data indicate that the recessive model of inheritance is not supported, as shown in the bottom of Table 9. The significant risk of one dose of the CA4 allele suggests that either the CA4 itself or the CA4 acting as a marker for a disease allele or haplotype is being inherited in a dominant fashion.

With identification of significant LD in *MSX1*, a search for causal mutations was undertaken using a diagnostic re-

Table 8
Distribution of triad types^a for *MSX1* gene variants among 162 sporadic orofacial cleft cases

Mating Type (MFC)	<i>MSX1</i> -CA ^a			
	Total	CLP	CLO	CPO
222	46	32	12	2
212	27	21	5	1
211	11	4	4	3
122	18	10	7	1
121	20	13	6	1
201	8	5	3	0
021	6	5	0	1
112	5	3	2	0
111	7	2	4	1
110	2	1	1	0
101	5	3	1	1
100	1	2	0	0
011	4	2	1	0
010	2	1	1	0
000	0	0	0	0

^aThe triad types were generated from two sets of *MSX1* CA variant repeats: 1 = 1, 2, 3, or 5 CAs and 2 = 4 CAs, as defined in Table 6.

quencing protocol that sequenced each of the two coding exons and a portion of the intervening sequence that is highly conserved between human and mouse. A full list of all the mutations found in this study can be found in Table 5.

Table 9
Relative Risk—Log Linear modeling analysis of proband’s *MSX1* CA 4 dose

		Relative risk	
		# <i>MSX1</i> CA4 alleles	
	0	1	2
Model parameters			
No assumption of HWE	1	0.89 < 2.43 < 6.63 <i>P</i> = 0.084	1.20 < 3.48 < 10.11 <i>P</i> = 0.022
Assume HWE	1	1.03 < 2.64 < 6.79 <i>P</i> = 0.044	1.29 < 3.60 < 10.06 <i>P</i> = 0.015
Assume HWE with recessive model	1	1	0.90 < 1.36 < 2.05 <i>P</i> = 0.14

For the log-linear *MSX1* CA relative risk assessment, there were 162 complete triads of data used and of these, there were 104 CL/P, 47 CLO, and 11 CPO patients (see supplemental data at <http://genetics.uiowa.edu/publications.html> for results by subgroup).

Missense mutations

Among the seven families recruited to this study with a positive family history of clefting, one of these families was found to have a missense mutation, P147Q, that was also found in two families from the sporadic group. The C to A transversion mutation found at position 440 of exon 1 changes the proline to glutamine at position 147 of the *MSX1* protein sequence. This result is shown along with the three families’ pedigrees in Fig. 1. The P147Q missense mutation was found in the affected child and the affected father, as well as in his aunt. The P147Q

allele is inherited in a dominant but incompletely penetrant pattern within this family. Interestingly, two family members who carry the P147Q mutation in this family also show some form of deafness, although only minimal data are available on this aspect of the phenotype. The father of the proband has been deaf since he was 10 years old. His son was born with CLP and was 3 years old at the time of ascertainment of all data. This father’s sister, who was found to also carry the P147Q mutation and give birth to an infant with a cleft, also has some form of unclassified deafness.

Family two, as shown in Fig. 1, was also found to harbor a P147Q mutation in addition to a novel mutation, G98E. Neither of these mutations was found in a previous mutation scan of over 1000 cases and controls.⁹ This included sequencing of 716 Asian control individuals for exon1 of *MSX1* (Table 2). Thus family 2 has one of each of these rare missense alleles in the parents: P147Q in the unaffected mother and G98E in the unaffected father. When both alleles are brought together in the proband, we find a cleft that may result from the combined effects of these two mutations. A third family was also identified with the same P147Q mutation found in the unaffected father and the infant daughter who has CPO.

The discovery of the second and third apparently unrelated families with the same highly conserved amino acid change, P147Q, suggests that either these families are closely related or this is a common population specific disease allele. Two of the families come from Qui Nhon Province, whereas family 2 was from Ben Tre Province several hundred miles to the south (see Fig. 1). First, Mendelian inheritance was checked by typing 4

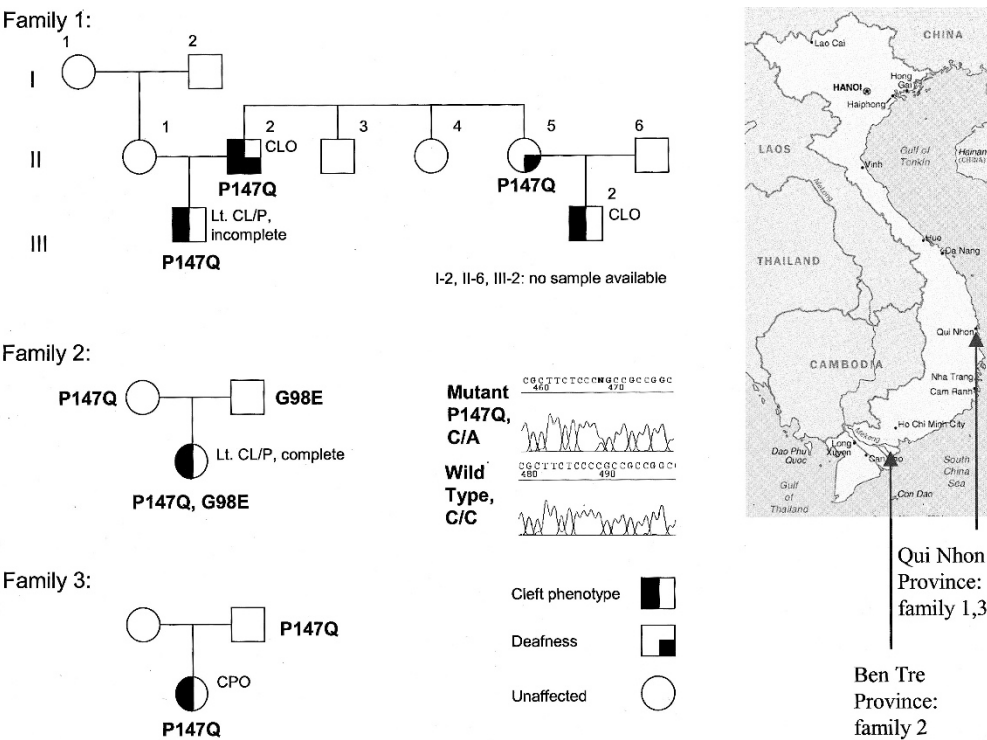


Fig. 1. Pedigrees of families with the *MSX1* missense variants. Map at the right shows the province of origin for these families.

microsatellite markers on 4 different chromosomes in all the members of these three families. All data from this work was consistent with proper Mendelian inheritance (data not shown). Next, haplotype analysis was performed within these families using TaqMan assays for common single nucleotide polymorphisms. The results of the haplotype analysis are shown in Table 10. From this data it is clear that all individuals carrying the P147Q mutation have AATTG4 bolded alleles in common and may have inherited this haplotype by descent. The same set of markers were typed in 100 unrelated Vietnamese controls as well as in the 175 case-triads. This haplotype was not disease associated in the case-triads (data not shown). The control haplotype frequencies are shown in Table 11. The AATTG4 haplotype was found in only 7.5% of the 200 control chromosomes. The chance of finding this relatively rare haplotype in three mutation families such as this is therefore approximately 0.6% ($1.00 \times 0.075 \times 0.075 = 0.0057$), suggesting that these families share a common ancestor and the allele is likely inherited by descent from a founder mutation.

DISCUSSION

The present report extends previous work on three candidate genes in Asian populations.^{5,6,9} Several previous studies support a role for the *MSX1* gene in the etiology of cleft lip and palate in humans. The mouse knockout of *MSX1*²⁰ has a cleft palate as well as dental and other craniofacial anomalies. A large family with a stop codon in the N-terminal portion of the protein cosegregates with the phenotype of cleft lip and palate or cleft palate alone as well as specific tooth agenesis.⁸ Several studies using association or transmission distortion have also

provided evidence for the involvement of *MSX1*.^{5,9,12,17,21} 4p16 deletions (Wolf-Hirschhorn syndrome) have a high frequency of CL/P.²²

This study reports transmission distortion of the 4 allele of a CA repeat found within a single intron of *MSX1* with nonsyndromic cleft lip and palate. This is the same allele associated with clefting in the multiple previously published studies as well as additional studies reported in abstract form. These prior positive results give strength to the present findings of significant overtransmission of the CA4, even though the biallelic TDT is not corrected for multiple comparisons. The present results then serve as confirmatory evidence for the involvement of *MSX1* in orofacial clefting in an independent population.

The continued involvement of the CA4 allele as the associated or overtransmitted allele in these studies might suggest that this allele itself is contributing to a predisposition to clefting in some families. One possible argument that could explain a potential disease causative role for the CA repeat relates to the existence of an *MSX1* antisense (AS) transcript. This antisense transcript appears to be a negative regulator of *MSX1* expression. It heteroduplexes with the *MSX1* sense (S) transcript to cause accelerated degradation of the *MSX1* sense mRNA.^{23,9} The *MSX1* CA repeats are found embedded within the sequence of the antisense transcript. It is conceivable that AS/S heteroduplexes of different lengths, as might be caused by different allelic numbers of *MSX1* CA repeats, could have an effect on the antisense binding to/degradation of the sense strand. However, such mechanisms will remain speculative until direct functional studies can be performed. Alternatively, the failure to identify a more recognizable etiologic mutation

Table 10
Haplotype data for P147Q family members

Family ID	Indiv ID	Missense variants ^a	Variants used for haplotyping ^b						CA repeats
		P147Q	G98E	snp1	snp2	snp3	snp4	snp5	
1	I-1			AG	AG	TT	GG	TT	44
	II-1			GG	GG	TT	GG	TT	44
	II-2	X/CLO + DF		AG	AG	TT	TG	GT	44
	II-3			AG	AG	TT	GG	TT	44
	II-4			AG	AG	TT	GG	TT	44
	II-5	X/DF		AG	AG	TT	TG	GT	44
	III-1	X/CLP		AG	AG	TT	TG	GT	44
2	I-1	X		AA	AA	TT	TT	GG	44
	I-2		X	AA	AG	TT	TG	GT	22
	II-1	X/CLP	X	AA	AA	TT	TT	GG	42
3	I-1			AA	AA	TT	TG	GT	42
	I-2	X		AA	AA	TT	TT	GG	44
	II-1	X/CPO		AA	AA	TT	TT	GG	42

^aX indicates presence of missense variant. CLO is cleft lip only, CLP is cleft lip with cleft palate, CPO is isolated cleft palate and DF is deafness.

^bPositions of the different variants used in this haplotype analysis are shown in Table 4.

Table 11
Top 90% of haplotypes from 100 unrelated control Vietnamese

Haplotype#	snpl	snp2	snp3	snp4	snp5	CA repeat	Frequency (#/total chromosomes)
1	A	A	T	T	T	4	56/200
2	A	A	T	G	G	2	53/200
3	G	G	T	T	T	4	45/200
4	A	A	T	T	G	4	15/200
5	A	G	T	T	T	4	11/200

than a tandem CA repeat despite extensive sequencing of the *MSX1* gene in a collection of over 800 individuals with non-syndromic cleft lip and palate, including many of Asian descent⁹ suggests that already identified variants within the gene may themselves be etiologic, perhaps within specific haplotypes. The possibility that sequence disruptions lie outside of the sequenced region in 5' or 3' regulatory regions can also not be dismissed. Nevertheless, it is still imperative to examine the genetic evidence for or against the CA4 allele being an actual disease mutation.

In this context, analysis of the effect of CA4 dose is relevant, because it is postulated above that heteroduplex length discrepancies might change sense processing. In previous analyses of the *MSX1* CA 4 dosage effects,²¹ the relative risk of disease increased with dose but not significantly so. However, Jugessur et al.²¹ found an almost two-fold increased risk among the CPO probands that were either heterozygous or homozygous for the variant allele, but again this dominant pattern was not statistically significant. Nevertheless, these results foreshadow the present significant results. In this Vietnamese cleft triad population, the *MSX1* CA 4 is inherited in a dominant fashion where the existence of one CA4 allele is enough for disease association. (Table 9).

It is most likely that the *MSX1* CA4 allele, being the most common allele in all populations studied to date, has picked up disease variants/mutations that are still in strong LD with the CA 4 allele. The search for these variants in regulatory domains both within the intron and found outside the transcriptome continues.

The sequencing analysis of the *MSX1* gene found two new amino acid variants, one in a family with deafness segregating with the variant allele. The *Msx1* null mouse²⁰ had an ear phenotype: although the incus and the stapes appeared normal, the processus brevis of the malleus fails to form in all homozygotes and is reduced in height overall. The malleus is a first pharyngeal arch derivative. In addition, the double knock out of murine *Msx1* and *Gsc* genes²⁴ revealed that although the phenotype was again restricted to the first arch derivatives, a more severe or additive phenotype was found in this double KO mice with a loss of the distal portions of the malleus. Recently,²⁵ it was shown that *Msx1* and *Msx2* have overlapping expression in the malleus. It was also shown that a transgenically rescued *Msx1* null mouse that was still missing the malleal

processus brevis could hear normally. Yet family 1 had two members who had some form of deafness that carried the P147Q allele. Thus either those members of family 1 had unrelated causes (gene and/or environmental) of hearing loss or the P147Q mutation might be acting in a dominant-negative fashion wherein the resulting phenotype is more severe.

Proline 147 is conserved in *MSX1* genes throughout mammalian and vertebrate evolution,⁹ (see supplemental data at <http://genetics.uiowa.edu/publications/peterj/img1.jpg>). It is found within a highly conserved block of amino acids just upstream of the homeodomain portion of the protein. This region has been called the extended homeodomain and there are multiple sequence motifs and potential functions that could be disrupted within this region. The strong conservation of the amino acid and the surrounding amino acids, along with the finding that the P147Q variant is segregating with the phenotype in multiple affected family members is highly suggestive that this variant is etiologic.

For the G98E mutation, there is less evidence for evolutionary conservation. However, by reference to the phylogenetic alignment of all the *MSX1* sequences in this region,⁹ all the amino acids at position 98 are nonpolar (e.g., Glycine or Alanine), whereas E (glutamine) is clearly negatively charged and polar. Thus there is a substantial change in amino acid class at position 98. A similar variant, G91D, was previously found only in a cleft case and the unaffected mother within a Filipino family. Neither of these mutations have been found in the screening of over 1000 control individuals of various ethnic backgrounds⁹ (Table 2).

The P147Q variant was found in three different, apparently unrelated families who share a relatively rare haplotype presumably inherited by descent. This variant may then represent a low frequency, population-specific, missense disease allele. In light of this discussion, it may be inherited as an incompletely penetrant, dominant allele. The presence of this population-specific disease variant justifies the continued search for causes of orofacial clefting in different populations around the world. As shown here for the common disease marker as well as for the individually important P147Q mutation, orofacial clefting is a heterogeneous complex disease with different genes and environmental factors impacting each population in a unique manner.

Of interest is the finding that the sporadic cases had more significant *P* values for both *MSX1* and *TGFB3* markers. This may indicate that common alleles at these particular markers are modifiers and are therefore influential in the sporadic group, whereas in the familial group other variants (such as P147Q) might act with major effect, without the need for additional modification from the associated alleles for the phenotype to be manifest.

The continued demonstration of the CA 4 allele of *MSX1* associated with nonsyndromic cleft lip and palate in this Asian population suggests that there are some common etiologies underlying nonsyndromic clefting across all population groups or at least Asians and Europeans. The genetic evidence for the involvement of *MSX1* does not eliminate environmental causes, but does suggest that a subset of patients can now be identified with a strong genetic etiology. A group of individuals lacking this 4 allele might then provide a more powerful resource for investigation of other genetic or environmental causes. Given the strong anecdotal evidence of a role for environmental exposures, particularly that secondary to dioxin, which was a widely used defoliant during the Vietnam war, it is important to distinguish identifiable causes from those that may have, as yet, undiscovered etiologies. Further, comparing the incidence of clefting and the causes of clefting between Vietnamese and other Asian populations might shed more light on the impact of specific environmental exposures as etiologies.

Finally, the demonstration that missense mutations can explain about 2% of the cases of nonsyndromic clefting in this population is consistent with one previous report.⁹ Because families with missense mutations appear to have higher penetrance alleles than those found at least empirically in families without identified *MSX1* mutations, we should begin consideration of *MSX1* sequencing in individuals and particularly in families with multiple affected members with clefts and/or dental anomalies outside the region of the cleft. This can be the first step in providing improved genetic counseling to at least a subset of families who currently have only broadly defined empiric recurrence risks. Further delineation of the full phenotypic spectrum and penetrance of *MSX1* mutations in families with clefts remains for future studies to define. In the end, knowledge of the specific genetic factors will also allow for a better determination of whether particular environmental covariates such as smoking or vitamin use might also play greater or lesser roles in a given individual. This will further refine our ability to not only diagnosis and predict cleft occurrences but to act in a more targeted way for its prevention.

ACKNOWLEDGMENTS

This work was supported by grants DE08559, DE13076, and DE10876. We are extremely grateful to the many patients, physicians, and volunteers in Vietnam who willingly contributed to these studies. We especially thank the AGU High-Tech Research Center Project. We also benefited from the technical assistance of Carla Nishimura, Bonnie Ludwig, and Sarah O'Brien.

References

1. Murray JC. Gene/environment causes of cleft lip and/or palate. *Clin Genet* 2002;61:248–256.
2. Romitti PA, Lidral AC, Munger RG, Daack-Hirsch S, Burns TL, Murray JC. Candidate genes for nonsyndromic cleft lip and palate and maternal cigarette smoking and alcohol consumption: evaluation of genotype-environment interactions from a population-based case-control study of orofacial clefts. *Teratology* 1999;59:39–50.
3. Mossey PA, Little J. Epidemiology of oral clefts: an international perspective. In: *Cleft Lip & Palate, From Origin to Treatment*. Wyszynski DF, editor. Oxford University Press, 2002:127–158.
4. Mitchell LE. Transforming growth factor alpha locus and nonsyndromic cleft lip with or without cleft palate: a reappraisal. *Genet Epidemiol* 1997;14:231–240.
5. Lidral AC, Romitti PA, Basart AM, Doetschman T, Leysens NJ, Daack-Hirsch S et al. Association of *MSX1* and *TGFB3* with nonsyndromic clefting in humans. *Am J Hum Genet* 1998;63:557–568.
6. Sato F, Natsume N, Machida J, Suzuki S, Kawai T. Association between Transforming Growth Factor Beta 3 and Cleft Lip and/or Palate in the Japanese Population. *Plast Reconstr Surg* 2001;107:1909–1910.
7. Beaty TH, Hetmanski JB, Zeiger JS, Fan YT, Liang KY, VanderKolk CA, McIntosh I. Testing candidate genes for non-syndromic oral clefts using a case-parent trio design. *Genet Epidemiol* 2002;22:1–11.
8. van den Boogaard MJ, Dorland M, Beemer FA, van Amstel HK. *MSX1* mutation is associated with orofacial clefting and tooth agenesis in humans. *Nat Genet* 2000;24:342–343. Erratum in: *Nat Genet* 2000;25:125.
9. Jezewski P, Vieira AR, Nishimura C, Ludwig B, Johnson M, O'Brien SE et al. Complete sequencing shows a role for *MSX1* in nonsyndromic cleft lip and palate. *J Med Genet* 2003;40:399–407.
10. Natsume N, Kawai T, Huynh L. Incidence of CLP in Vietnam (Letter to the Editor). *Cleft Palate Craniofac J* 1998;35:183.
11. Shi M, Caprau D, Dagle J, Christiansen L, Christensen K, Murray JC. Application of kinetic polymerase chain reaction and molecular beacon assays to pooled analyses and high-throughput genotyping for candidate genes. *Birth Defects Res Part A Clin Mol Teratol* 2004;70:65–74.
12. Fallin MD, Hetmanski JB, Park J, Scott AF, Ingersoll R, Fuernkranz HA et al. Family-based analysis of *msx1* haplotypes for association with oral cleft. *Genet Epidemiol* 2003;25:168–175.
13. Horvath S, Xu X, Laird N. The family based association test method: strategies for studying general genotype-phenotype associations. *Eur J Hum Gen* 2001;9:301–306.
14. Weinberg CR, Wilcox AJ, Lie RT. A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. *Am J Hum Genet* 1998;62:969–978.
15. Ardinger HH, Buetow KH, Bell GI, Bardach J, VanDemark DR, Murray JC. Association of genetic variation of the transforming growth factor- α gene with cleft lip and palate. *Am J Hum Genet* 1989;45:348–353.
16. Lidral AC, Murray JC, Buetow KH, Basart AM, Scheerer H, Shiang R et al. Studies of the candidate genes *TGFB2*, *MSX1*, *TGFA*, and *TGFB3* in the etiology of cleft lip and palate in the Philippines. *Cleft Palate Craniofac J* 1997;34:1–6.
17. Padanilam BJ, Stadler HS, Mills KA, McLeod LB, Solursh M, Lee B et al. Characterization of the human *HOX 7* cDNA and identification of polymorphic markers. *Hum Mol Genet* 1992;1:407–410.
18. Beaty TH, Wang H, Hetmanski JB, Fan YT, Zeiger JS, Liang KY et al. A case-control study of nonsyndromic oral clefts in Maryland. *Ann Epidemiol* 2001;11:434–442.
19. Vieira AR, Orioli IM, Castilla EE, Cooper ME, Marazita ML, Murray JC. *MSX1* and *TGFB3* contribute to clefting in South America. *J Dent Res* 2003;82:289–292.
20. Satokata I, Maas R. *Mx1* deficient mice exhibit cleft palate and abnormalities of craniofacial and tooth development. *Nat Genet* 1994;6:348–356.
21. Jugessur A, Lie RT, Wilcox AJ, Murray JC, Taylor JA, Saugstad OD et al. Variants of developmental genes (*TGFA*, *TGFB3*, and *MSX1*) and their associations with orofacial clefts: a case-parent triad analysis. *Genet Epidemiol* 2003;24:230–239.
22. Battaglia A, Carey JC, Cederholm P, Viskochil DH, Brothman AR, Galasso C. Natural history of Wolf-Hirschhorn syndrome: experience with 15 cases. *Pediatrics* 1999;103:830–836.
23. Blin-Wakkach C, Lezot F, Ghoul-Mazgar S, Hotton D, Monteiro S, Teillac C et al. Endogenous *Mx1* antisense transcript: in vivo and in vitro evidences, structure, and potential involvement in skeleton development in mammals. *Proc Natl Acad Sci U S A* 2001;98:7336–7341.
24. Kuratani S, Satokata I, Blum M, Komatsu Y, Haraguchi R, Nakamura S et al. Middle ear defects associated with the double knock out mutation of murine gooseoid and *Mx1* genes. *Cell Mol Biol (Noisy-le-grand)* 1999;45:589–599.
25. Zhang Z, Zhang X, Avniel WA, Song Y, Jones SM, Jones TA et al. Malleal process brevis is dispensable for normal hearing in mice. *Dev Dyn* 2003;227:69–77.
26. den Dunnen JT, Antonarakis SE. Nomenclature for the description of human sequence variations. *Hum Genet* 2001;109:121–124.