Brief Genetics Report

A Comprehensive, Statistically Powered Analysis of *GAD2* in Type 1 Diabetes

Gillian C.L. Johnson,[1] Felicity Payne,[1] Sarah Nutland,[1] Helen Stevens,[1] Eva Tuomilehto-Wolf,[2] Jaakko Tuomilehto,[2] and John A. Todd[1]

GAD2 maps to chromosome 10p11.23 and encodes the 65-kDa isoform of GAD65, a major autoantigen in type 1 diabetes. The genetic variation that influences expression of preproinsulin mRNA, encoding another major autoantigen in type 1 diabetes, has already been shown to be genetically associated with disease. Previous reports that have assessed the association of *GAD2* with type 1 diabetes have not used a dense map of markers surrounding the gene and have relied on very small clinical sample sizes. Consequently, no definite conclusions can be drawn from their negative results. We have therefore systematically searched all exons, the 3′ untranslated region (UTR), the 5′ UTR, and the 5′ upstream region of *GAD2*, for polymorphisms in 32 white European individuals. We have genotyped these polymorphisms in a maximum of 472 U.K. type 1 diabetic affected sib pair families exhibiting linkage to type 1 diabetes on chromosome 10p and have tested both single variants and haplotypes in the *GAD2* region for association with disease. We subsequently followed up our results by genotyping a subset of these single-nucleotide polymorphisms in a maximum of 873 Finnish families with at least one affected child. Our results suggest that *GAD2* does not play a major role in type 1 diabetes in these two European populations. *Diabetes* 51:2866–2870, 2002

G AD65 has been identified as a major autoantigen in type 1 diabetes, with autoantibodies against GAD65 present in the sera of 50–80% of newly diagnosed type 1 diabetic patients as compared with ~2% of the general population (1). It is still unclear whether immune responses against GAD are directly involved in the destruction of the islet β-cells or are merely a consequence of other pathogenic mechanisms that cause type 1 diabetes. However, in NOD mice, diabetes is characterized by early spontaneous T-cell reactivity

to GAD, and prevention of this anti-GAD immunity results in the prevention of insulitis (β-cell autoimmunity) and diabetes (2–4). Although such compelling evidence for a primary role for GAD in human type 1 diabetes pathology has not been demonstrated, GAD autoantibodies can be detected several years before the onset of human disease (5), suggesting that autoimmunity to GAD could play a major role in the preclinical phase of type 1 diabetes (6).

Preproinsulin and GAD65, both strong candidates for primary type 1 diabetes autoantigens, are expressed in the thymus, where they can influence T-cell tolerance (7,8). Indeed, there is evidence to suggest that variation in thymic preproinsulin mRNA expression is associated with genetic predisposition to type 1 diabetes, encoded by a variable number tandem repeat locus 5′ of the insulin gene (*IDDM2*) (7,8), and it is possible that this could also be true for GAD65. GAD65 is encoded by *GAD2*, which maps to chromosome 10p11.23 (9,10). In addition to being a strong functional candidate, *GAD2* is also a strong positional candidate for involvement in disease susceptibility because chromosome 10p11.2 shows evidence of linkage to type 1 diabetes in U.K. sib-pairs (11,12). Three previous studies have specifically attempted to address whether *GAD2* is genetically associated with type 1 diabetes (12–14). However, none of these studies used a dense enough map of markers nor a large enough sample size to stand any reasonable chance of detecting a disease association, should one exist. Indeed, Reed et al. (12) recognized that only a high-resolution linkage disequilibrium (LD) map of single-nucleotide polymorphisms (SNPs) within the gene could comprehensively exclude a minor role for *GAD2* in type 1 diabetes predisposition. Consequently, no definite conclusions can be drawn from these studies (12,13).

The entire contiguous genomic sequence of *GAD2* can be found on a clone from the RPCI-13 BAC library, RP13-8106: AL389927 (available at www.sanger.ac.uk). The coding regions of *GAD2* are organized into 16 exons spanning ~88 kb, and the sequence and genomic structure of rat, murine, and human GAD2 genes appears to be highly conserved (15–18). Indeed, comparing the noncoding 5′ flanking sequence of human *GAD2* with upstream sequences in rat *Gad2* (AF090195) reveals specific regions of high sequence similarity (>90%) in a region that has previously been suggested to contain a major TATA-less promoter that controls gene expression (17). In addition, sequence comparison between noncoding 5′ flanking regions of mouse and rat *Gad2* also show high similarity
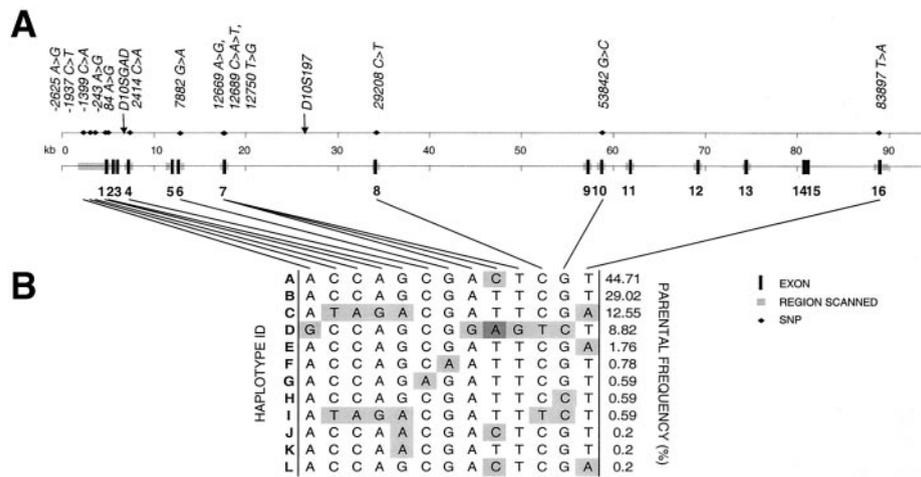
FIG. 1. *A*: Polymorphisms detected and genotyped at *GAD2*. *B*: U.K. parental haplotype reconstruction using 13 SNPs spanning the GAD2 gene.

(>90%), again suggesting the presence of conserved regulatory elements controlling expression of the gene (16,18).

To evaluate whether variants in *GAD2* influence susceptibility to type 1 diabetes, we scanned all coding sequences, most of the 3′ untranslated region (UTR), and 2.8 kb of the 5′ flanking region for polymorphisms (including the 5′ UTR and all regions showing similarity to 5′ flanking sequences in rat and murine *Gad2*) in 32 individuals. We detected 22 polymorphisms in total, 3 of which were in coding regions, and each caused a nonsynonymous change in the amino acid sequence of the protein. In addition to the 3 cSNPs detected in *GAD2*, we also genotyped 10 additional noncoding polymorphisms in a maximum of 472 multiplex families from the U.K. already showing evidence of linkage to disease in this region (11,12). This dataset provided >99.9% power to detect an effect equivalent to that seen at *IDDM2* (relative risk [RR] = 2, disease allele frequency = 0.7), with $P = 0.05$ and ~77% power to detect an effect at $P = 0.05$ for a disease allele with 5% frequency and an RR of 1.5. We selected the noncoding SNPs on the basis of their patterns of heterozygosity in the original screening panel of 32 individuals. As specific groups of SNPs were exclusively heterozygous in the same individuals, it is likely they all marked the same ancestral haplotype and therefore would be redundant in a first-pass association study. Therefore, we only chose a selection from each group to genotype in the diabetic family panel in the first instance. We also genotyped all other SNPs that were found to be heterozygous in at least three individuals in our screening panel. Our SNP map at *GAD2* therefore extended over 86.5 kb, from 2,625 bp upstream of the *GAD2* initiation codon to 91 bp before the start of the last exon (Fig. 1).

We observed 12 haplotypes when reconstructing parental haplotypes from all SNP markers (Fig. 1). A feature of the genomic architecture observed in the *GAD2* region is that SNP allele frequency is a good indicator of LD. Indeed, SNPs with very similar allele frequencies are in almost absolute LD and therefore mark certain haplotypes. For example, five SNPs whose minor allele frequency is ~8% (−2,625 A→G, 12,669 A→G, 12,750 T→G, 29,208 C→T, and 53,842 G→C) are mainly present on just one haplotype, haplotype D, and those with a minor allele frequency of ~15% (−1,937 C→T, −1,399 C→A, −243 A→G, and 84

A→G) are mainly present on another haplotype, haplotype C. Consequently, analyzing just two haplotype-tagging SNPs (19), 12,689 C→A→T, the SNP that segregates three alleles, and one of −1,937 C→T, −1,399 C→A, −243 A→G, or 84 A→G, can capture the majority of the variation that we observed in this 87-kb region (Fig. 1). Indeed, SNPs marking haplotype D and haplotype C represent the groups of polymorphisms that were heterozygous in the same individuals in our screening panel, confirming that a limited amount of additional information would have been gained if the remaining markers in these groups had been genotyped. We tested all individual SNPs for association with type 1 diabetes using the transmission-disequilibrium test (TDT) (Table 1). The group of five SNPs, mainly present on haplotype D, including both the 5′ flanking SNP, −2,625 A→G, which maps within a region of high sequence similarity (>90%) to a 5′ flanking region in rat *Gad2*, and also a SNP in exon 10 of *GAD2*, which causes a Gly-Glu amino acid change in the mature protein, showed a slight deviation from the expected Mendelian rate of transmission ($P < 0.05$). Analyzing the four most

TABLE 1
Association analysis of SNPs in the *GAD2* region in a maximum of 472 diabetic multiplex families originating from the U.K.

Marker	Allele	Transmitted	Nontransmitted	P^*
−2,625 A>G	A	127	91	0.02
−1,937 C>T	C	242	230	0.6
−1,399 C>A	A	186	191	0.82
−243 A>G	A	175	167	0.69
84 A>G	A	145	136	0.64
2,414 C>A	A	12	17	0.43
7,882 G>A	A	11	13	0.67
12,669 A>G	A	114	76	0.01
12,689				
C>A>T	A	80	108	0.04
	C	369	339	0.26
	T	376	378	0.95
12,750 T>G	G	77	117	0.01
29,208 C>T	C	100	78	0.11
53,842 G>C	C	134	169	0.05
83,897 T>A	A	172	200	0.2

*Robust variance estimators were used in the calculation of P values to ensure a valid test of association.

TABLE 2
Comparison of haplotypes observed in the *GAD2* region in the U.K. and Finnish populations

	−2,625 A>G	−1,937 C>T	84 A>G	12,669 A>G	12,689 C>A>T	12,750 T>G	83,897 T>A	Finnish parental frequency	U.K. parental frequency
a	A	C	G	A	C	T	T	42.56	43.92
b	A	C	G	A	T	T	T	37.66	31
c	A	T	A	A	T	T	A	9.97	13.53
d	G	C	G	G	A	G	T	6.17	8.66
e	A	C	G	A	T	T	A	1.27	1.82
f	A	T	A	A	T	T	T	0.95	0.61
g	A	C	G	A	C	T	A	0.79	0.15
h	A	C	G	A	T	T	T	0.16	—
i	A	C	G	A	T	T	A	0.16	—
j	A	T	A	A	C	T	A	0.16	—
k	G	C	G	G	C	G	T	0.16	—
l	A	C	A	A	C	T	T	—	0.15
m	A	C	A	A	T	T	T	—	0.15

common haplotypes captured by the two SNPs, 12,689 C→A→T and 84 A→G, revealed weak evidence for a protective effect of haplotype D (41.4% transmission, $P = 0.045$). No other haplotypes analyzed showed any significant deviations from Mendelian transmission to affected offspring. The two SNPs defining the most common SNP haplotypes were also analyzed in conjunction with two microsatellite markers, *D10S197* and GAD (13), in the region (Fig. 1). None of the four SNP/microsatellite haplotypes with a parental frequency >5% showed any evidence of association to disease (data not shown).

Although no additional U.K. families were available to extend the weak evidence for the association of SNPs on haplotype D with type 1 diabetes, 873 simplex families were available from Finland, the country with the highest incidence of type 1 diabetes (20). Initially, we genotyped seven of the *GAD2* polymorphisms (−2,625 A→G, −1,937 C→T, 84 A→G, 12,669 A→G, 12,689 C→A→T, 12,750 T→G, and 83,897 T→A) in 224 families to evaluate LD/haplotype patterns of these markers in the Finnish population. The four most common haplotypes comprising these six markers were identical in both populations (Table 2). Therefore, it seemed reasonable to assume that if a variant on haplotype D (with a frequency of ~8% in the U.K.) was causing susceptibility to type 1 diabetes in the U.K., it might also have the same effect in Finland. As the three SNPs that mark haplotype D in the U.K. were in absolute LD in the Finnish families studied, the remaining families were genotyped with only the −2,625 A→G polymorphism to evaluate association of this haplotype/SNP in Finnish diabetic individuals. In a total of 814 fully genotyped families, this SNP showed no evidence of association to type 1 diabetes using the TDT. Assuming the putative disease haplotype is having the same effect in both populations, and given the frequency of the SNP observed in the Finnish parental population (6.92%), these additional families should provide 77.8% power to detect a disease allele at $P = 0.05$ (assuming a multiplicative model [21]). Because our replication study was therefore reasonably powered, the initial weak result obtained in our U.K. families is likely to represent a false positive.

When a disease is known to be clinically heterogeneous, it may be beneficial to stratify the dataset on the basis of relevant predefined criteria (11,13). The HLA complex on chromosome 6p21.31 (*IDDM1*) contains genes encoding cell surface molecules that display peptides for immune recognition, with *HLA-DRB1*03* (*HLA-DR3*) and *HLA-DRB1*04* (*HLA-DR4*) haplotypes showing the greatest association to type 1 diabetes in European populations (22). GAD is a major autoantigen in type 1 diabetes, and peptide binding to class II molecules could be allotype specific. Indeed, it has been suggested that the early appearance of islet autoantibodies is common for DR3/4 heterozygous and DR4/4 homozygous offspring of type 1 diabetic parents (23) and that antibodies to GAD are specifically associated with DR3 alleles in siblings of type 1 diabetic subjects (24).

Offspring were therefore selected on the basis of having at least one copy of the high-risk HLA-DR3 or HLA-DR4 haplotypes to test whether the weak association seen with SNP 2,625 A→G in the original U.K. population was influenced by specific interactions with *IDDM1*. There was, however, no evidence for HLA-specific heterogeneity in the distribution of the 2625A→G*G allele in unrelated patients from either the U.K. or Finland (Table 3).

Our polymorphism search at GAD2 included all coding sequences and a 5′ flanking region that likely plays a role in controlling expression of the gene (17). Moreover, as the mean level of LD between pairs of loci within the gene was considerable ($D'_{mean} = 0.86$), it is possible that we would have also been able to detect, indirectly, the effects of any unidentified functional intronic variants. Combined with the large number of affected offspring ($n = 914$) studied in our first-pass scan, this should have allowed us to detect an effect equivalent to that of the insulin gene/*IDDM2* locus (RR = 2, disease allele frequency = 0.7) or larger at $P < 0.05$ with >99.9% power. However, we only found very modest evidence for a disease association with a single haplotype ($P = 0.045$), a finding that was not replicated in a second large dataset comprising 814 families with at least one affected offspring. Therefore, the data from this first comprehensive evaluation of the role of GAD2 in type 1 diabetes suggest that genetic variation within this gene is unlikely to contribute significantly to the development of the disease. It is possible that a functional variant of a GAD2 transcription control element, which is out of LD with the SNPs we have typed, exists some distance from the GAD2 structural gene.

TABLE 3
Proportion of chromosomes from unrelated patients with a given HLA phenotype carrying the $-26,25A>G*G$ allele.

	DR3+	DR3−	DR4+	DR4−	DR3/4	Non-DR3/4	P
U.K.	29/498 (5.8)	14/284 (4.9)	33/610 (5.4)	10/172 (5.8)	22/352 (6.2)	21/430 (4.9)	NS
Finland	24/308 (7.8)	29/432 (6.7)	36/530 (6.8)	17/210 (6.1)	12/188 (6.4)	41/552 (7.4)	NS

Data are n (%).

However, our results also suggest that present resources would best be directed at other candidate genes or regions within the putative *IDDM10* locus.

RESEARCH DESIGN AND METHODS

Subjects. We genotyped polymorphisms at *GAD2* in a maximum of 472 multiplex families originating from the U.K. in which both sibs were diagnosed with type 1 diabetes (the Diabetes U.K. Warren 1 repository). We also genotyped markers in a maximum of 873 Finnish families in which at least one child had been diagnosed with type 1 diabetes.

Polymorphism detection. Polymorphisms in *GAD2* were detected by screening 18 parents of type 1 diabetic sibs, 4 type 1 diabetics subjects, and 10 white control subjects not ascertained for disease status using DHPLC (WAVE). All individuals screened were white and born in the U.K. All regions screened were within 5 kb of the 5′UTR and 3 kb of the 3′UTR of the gene.

Genotyping. All polymorphisms were genotyped using Invader, as described by Mein et al. (25). Full details of all SNPs are available at our website (http://www-gene.cimr.cam.ac.uk/todd/).

Statistical analysis. Parental haplotypes were reconstructed from pedigree data using software available from Frank Dudbridge (http://www.hgmp.mrc.ac.uk/~fdudbrid/software).

The TDT was performed with STATA using the genassoc package available from David Clayton's website (http://www-gene.cimr.cam.ac.uk/clayton/software/). To ensure a valid test of association in multiplex families, robust variance estimators were used in the calculation of P values. TDT on multilocus haplotypes was performed using the software of Frank Dudbridge.

Contingency table analysis was used to compare allele frequencies of $-2,625\ A{\rightarrow}G$ between groups of diabetic offspring stratified by HLA genotype. Only one diabetic subjects was chosen from each family. The P values shown are not corrected for multiple tests.

REFERENCES

1. Atkinson MA, Maclaren NK, Scharp DW, Lacy PE, Riley WJ: 64,000 Mr autoantibodies as predictors of insulin-dependent diabetes. *Lancet* 335:1357–1360, 1990
2. Kaufman DL, Clare-Salzler M, Tian J, Forsthuber T, Ting GS, Robinson P, Atkinson MA, Sercarz EE, Tobin AJ, Lehmann PV: Spontaneous loss of T-cell tolerance to glutamic acid decarboxylase in murine insulin-dependent diabetes. *Nature* 366:69–72, 1993
3. Tisch R, Yang XD, Singer SM, Liblau RS, Fugger L, McDevitt HO: Immune response to glutamic acid decarboxylase correlates with insulitis in non-obese diabetic mice. *Nature* 366:72–75, 1993
4. Yoon JW, Yoon CS, Lim HW, Huang QQ, Kang Y, Pyun KH, Hirasawa K, Sherwin RS, Jun HS: Control of autoimmune diabetes in NOD mice by GAD expression or suppression in beta cells. *Science* 284:1183–1187, 1999
5. Tuomilehto J, Zimmet P, Mackay IR, Koskela P, Vidgren G, Toivanen L, Tuomilehto-Wolf E, Kohtamaki K, Stengard J, Rowley MJ: Antibodies to glutamic acid decarboxylase as predictors of insulin-dependent diabetes mellitus before clinical onset of disease. *Lancet* 343:1383–1385, 1994
6. von Boehmer H, Sarukhan A: GAD, a single autoantigen for diabetes. *Science* 284:1135–1137, 1999
7. Pugliese A, Zeller M, Fernandez Jr A, Zalcberg L, Bartlett R, Ricordi C, Pietropaolo M, Eisenbarth G, Bennett S, Patel D: The insulin gene is transcribed in the human thymus and transcription levles correlate with allelic variation at the *INS* VNTR-*IDDM2* susceptibility locus for type 1 diabetes. *Nat Genet* 15:293–297, 1997
8. Vafiadis P, Bennet S, Todd J, Nadeau J, Grabs R, Goodyer C, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by *INS* VNTR alleles at the *IDDM2* locus. *Nat Genet* 15:289–292, 1997
9. Karlsen AE, Hagopian WA, Grubin CE, Dube S, Disteche CM, Adler DA, Barmeier H, Mathewes S, Grant FJ, Foster D, et al: Cloning and primary structure of a human islet isoform of glutamic acid decarboxylase from chromosome 10. *Proc Natl Acad Sci U S A* 88:8337–8341, 1991
10. Bu DF, Erlander MG, Hitz BC, Tillakaratne NJ, Kaufman DL, Wagner-McPherson CB, Evans GA, Tobin AJ: Two human glutamate decarboxylases, 65-kDa GAD and 67-kDa GAD, are each encoded by a single gene. *Proc Natl Acad Sci U S A* 89:2115–2119, 1992
11. Mein CA, Esposito L, Dunn MG, Johnson GC, Timms AE, Goy JV, Smith AN, Sebag-Montefiore L, Merriman ME, Wilson AJ, Todd JA: A search for type 1 diabetes susceptibility genes in families from the United Kingdom. *Nat Genet* 19:297–300, 1998
12. Reed P, Cucca F, Jenkins S, Merriman M, Wilson A, McKinney P, Bosi E, Joner G, Ronningen K, Thorsby E, Undlien D, Merriman T, Barnett A, Bain S, Todd J: Evidence for a type 1 diabetes susceptibility locus (IDDM10) on human chromosome 10p11–q11. *Hum Mol Genet* 6:1011–1016, 1997
13. Wapelhorst B, Bell GI, Risch N, Spielman RS, Concannon P: Linkage and association studies in insulin-dependent diabetes with a new dinucleotide repeat polymorphism at the GAD65 locus. *Autoimmunity* 21:127–130, 1995
14. Rambrand T, Pociot F, Ronningen K, Nerup J, Michelsen BK: Genetic markers for glutamic acid decarboxylase do not predict insulin-dependent diabetes mellitus in pairs of affected siblings: the Danish Study Group of Diabetes in Childhood. *Hum Genet* 99:177–185, 1997
15. Bu DF, Tobin AJ: The exon-intron organization of the genes (GAD1 and GAD2) encoding two human glutamate decarboxylases (GAD67 and GAD65) suggests that they derive from a common ancestral GAD. *Genomics* 21:222–228, 1994
16. Pinal CS, Cortessis V, Tobin AJ: Multiple elements regulate GAD65 transcription. *Dev Neurosci* 19:465–475, 1997
17. Skak K, Michelsen BK: The TATA-less rat GAD65 promoter can be activated by Sp1 through non-consensus elements. *Gene* 236:231–241, 1999
18. Makinae K, Kobayashi T, Shinkawa H, Sakagami H, Kondo H, Tashiro F, Miyazaki J, Obata K, Tamura S, Yanagawa Y: Structure of the mouse glutamate decarboxylase 65 gene and its promoter: preferential expression of its promoter in the GABAergic neurons of transgenic mice. *J Neurochem* 75:1429–1437, 2000
19. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA: Haplotype tagging for the identification of common disease genes. *Nat Genet* 29:233–237, 2001
20. Tuomilehto J, Karvonen M, Pitkaniemi J, Virtala E, Kohtamaki K, Toivanen L, Tuomilehto-Wolf E: Record-high incidence of type I (insulin-dependent) diabetes mellitus in Finnish children: the Finnish Childhood Type I Diabetes Registry Group. *Diabetologia* 42:655–660, 1999
21. Mitchell LE: Relationship between case-control studies and the transmission/disequilibrium test. *Genet Epidemiol* 19:193–201, 2000
22. Herr M, Dudbridge F, Zavattari P, Cucca F, Guja C, March R, Campbell RD, Barnett AH, Bain SC, Todd JA, Koeleman BP: Evaluation of fine mapping strategies for a multifactorial disease locus: systematic linkage and association analysis of IDDM1 in the HLA region on chromosome 6p21. *Hum Mol Genet* 9:1291–1301, 2000
23. Schenker M, Hummel M, Ferber K, Walter M, Keller E, Albert ED, Janka HU, Kastendiek C, Sorger M, Louwen F, Ziegler AG: Early expression and high prevalence of islet autoantibodies for DR3/4 heterozygous and DR4/4 homozygous offspring of parents with type I diabetes: the German BABY-DIAB study. *Diabetologia* 42:671–677, 1999

24. Kulmala P, Savola K, Reijonen H, Veijola R, Vahasalo P, Karjalainen J, Tuomilehto-Wolf E, Ilonen J, Tuomilehto J, Akerblom HK, Knip M: Genetic markers, humoral autoimmunity, and prediction of type 1 diabetes in siblings of affected children: Childhood Diabetes in Finland Study Group. *Diabetes* 49:48–58, 2000

25. Mein CA, Barratt BJ, Dunn MG, Siegmund T, Smith AN, Esposito L, Nutland S, Stevens HE, Wilson AJ, Phillips MS, Jarvis N, Law S, de Arruda M, Todd JA: Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. *Genome Res* 10:330–343, 2000