Blood Cells, Molecules and Diseases 55 (2015) 320–327

Contents lists available at ScienceDirect

Blood Cells, Molecules and Diseases

journal homepage: www.elsevier.com/locate/bcmd

A founder effect for p47^{phox} *Trp193Ter* chronic granulomatous disease in Kavkazi Jews

Martin de Boer [a,1], Shay Tzur [b,1], Karin van Leeuwen [a,1], Paula C.D. Dencher [a], Karl Skorecki [b], Baruch Wolach [c], Ronit Gavrieli [c], Ivane Nasidze [d,2], Mark Stoneking [d], Michael W.T. Tanck [e], Dirk Roos [a,*]

[a] Sanquin Research, and Landsteiner Laboratory, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands
[b] Dept. of Nephrology and Molecular Medicine, Rappaport Faculty of Medicine and Research Institute, Technion, and Rambam Medical Center, Haifa, Israel
[c] Dept. of Pediatrics and Laboratory for Leukocyte Function, Meir Medical Center, Kfar Saba, Israel
[d] Dept. of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
[e] Dept. of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Submitted 14 July 2015
Accepted 15 July 2015
Available online 23 July 2015

Editor: Mohandas Narla

Keywords:
Chronic granulomatous disease
NADPH oxidase
p47^{phox}
NCF1
Kavkazi Jews
Founder effect

ABSTRACT

Chronic granulomatous disease (CGD) is a rare congenital immune deficiency caused by mutations in any of the five genes encoding NADPH oxidase subunits. One of these genes is *NCF1*, encoding the p47^{phox} protein. A group of 39 patients, 14 of whom are of Kavkazi Jewish descent, was investigated for a founder effect for the mutation c.579G > A (*p.Trp193Ter*) in *NCF1*. We analyzed various genetic markers in the *NCF1* region, including two single nucleotide polymorphisms (SNPs) in *NCF1* and two short tandem repeats (STRs) located near *NCF1*. Most patients were homozygous for the c.579G > A mutation, but three patients were hemizygotes, with a deletion of *NCF1* on the other allele, and three patients were compound heterozygotes with another mutation in *NCF1*. All Kavkazi Jewish patients had a c.295G_c.345T SNP combination in *NCF1* and shared a common number of repeats in STR3. In addition, 90% of the Kavkazi Jewish patients shared a common number of repeats in STR1. This uniformity indicates that the c.579G > A mutation in *NCF1* was introduced some 1200–2300 years ago in the Kavkazi Jewish population. Variation amongst the other investigated populations from the Middle East indicates that this mutation exists in these non-Kavkazi populations already for more than 5000 years.

1. Introduction

Chronic granulomatous disease (CGD) is a rare (1:250,000 births) congenital disorder in which phagocytic leukocytes cannot generate superoxide (O_2^-) and other microbicidal reactive oxygen species owing to mutations in one of the five components of the O_2^- generating NADPH oxidase enzyme complex. The most common autosomal subtype of CGD is caused by a frameshift mutation in exon 2 (c.75_76delGT) of neutrophil cytosol factor 1 (*NCF1*), encoding the NADPH oxidase subunit p47^{phox} [1]. This mutation is the result of unequal sister chromatid exchange of *NCF1* and one of its two pseudogenes at chromosome 7.11.23 [ref. 2, 3, 4].

However, in a select group of p47^{phox}-deficient patients, CGD is caused by the mutation c.579G > A (*p.Trp193Ter*) in exon 7 of *NCF1*, which leads to the premature termination of p47^{phox} protein synthesis [1,5,6]. We studied a group of 39 of such patients, of whom 14 are of Kavkazi Jewish descent from 12 different families. The Jewish community of the Caucasus, also known as Kavkazi Jews or Mountain Jews, is believed to have been established during the 8th century C.E. in the region corresponding to Dagestan and the current state of Azerbaijan, as a result of a movement of Jews from Iran (Fig. 1) [ref. 7]. In recent years, many descendants from these people have returned to Israel. We hypothesized that these Kavkazi Jewish patients may share a common ancestor with the c.579G > A mutation. Similar founder effects for a number of lysosomal storage diseases in the Ashkenazi Jews have been recognized [8]. The remaining 25 CGD patients with the c.579G > A mutation in *NCF1* from 17 different families are also from the Middle East and South-East Europe (Fig. 1), but, with the exception of one family, of non-Jewish descent (Table 1). Because their origin is from a much wider geographical territory than that of the Kavkazi Jewish patients, and because they share the common cause of disease, they serve as an important comparison group. Similar c.579G > A CGD patients from other geographic regions have not been described and are not known to us.

Microsatellites such as short tandem repeats (STRs) can provide valuable phylogenetic insight into the evolution of a disease [9]. Their highly polymorphic nature distinguishes microsatellites from other DNA sequences, thereby serving as an informative category of genetic

* Corresponding author at: Sanquin Research, Plesmanlaan 125, 1066CX Amsterdam, The Netherlands.
E-mail address: d.roos@sanquin.nl (D. Roos).
[1] These authors contributed equally to this investigation.
[2] Deceased.

http://dx.doi.org/10.1016/j.bcmd.2015.07.014
1079-9796/

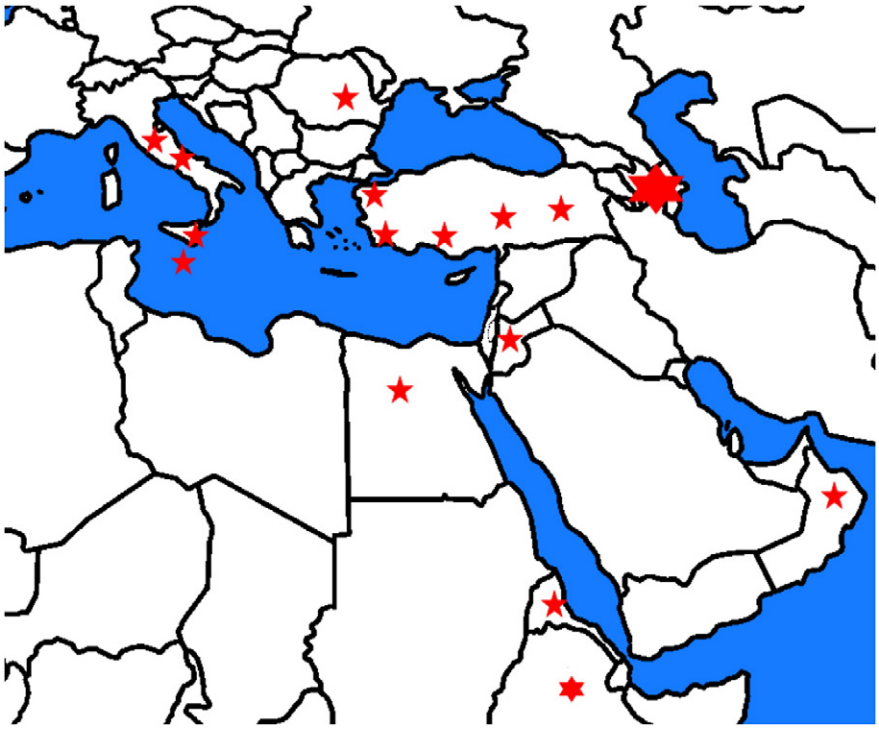

Fig. 1. Geographical distribution of the *NCF1* c.579G > A mutation. Six-pointed stars: Jewish patients. Five-pointed stars: non-Jewish patients.

markers for population genetics and forensic studies [10–14]. Other important genetic markers are single nucleotide polymorphisms (SNPs). Both SNPs and STRs are known to change and mutate over time. The overall variation in the categories of markers and their combinations reflect rates of mutation (single nucleotide substitutions for SNPs and number of repeats for STRs) as well as recombination exchange of sister chromatid material during meiosis. Since recombination occurs during oogenesis, and oocyte chromatids lie dormant until ovulation, this makes for a higher recombination rate in oocytes than in spermatozoa. Conversely, mutation tends to take place more often and with increasing paternal age in males, as spermatozoa are continuously being produced. There are also other factors that determine the rates at which recombination and mutation take place. In addition to paternal and maternal differences, a longer region of DNA over which repeats occur is more likely to experience recombination than shorter lengths, especially if this is paired with a higher number of repeats [14]. The location of a microsatellite is also of influence, occurring twice as often at the ends of chromosome arms as near the centromere [15,16]. The distribution of mutation rates can also differ across ethnic populations [14]. Replication slippage has been considered as the mechanism responsible for new mutations in microsatellites [17]. Moreover, the same number of repeats at a given single STR can occur through multiple different phylogenetic pathways that do not necessarily indicate shared ancestry, whereas such convergent evolution for a given SNP is much less common. Nonetheless, mutation and recombination rates have been determined that can be applied to calculate the approximate introduction of genetic variation.

Two STRs (STR1 and STR3) have been identified upstream of *NCF1* (Fig. 2). STR1 is located about 2 Mb upstream of *NCF1*. The recombination rate over a distance of 2 Mb has been determined as being 2% per generation (25 years), and the estimated mutation rate is 0.2% [10]. This gives a combined rate of change of 2.2% per generation. Therefore, a change in STR1, be it a mutation or recombination, can occur approximately every 1136 years ([100/2.2]*25). Based on the calculations mentioned above, the recombination rate of STR3 (about 0.25 Mb upstream of *NCF1*) is 0.25% per generation, and the mutation rate of STR3 is estimated to be the same as STR1, leading to a combined recombination and mutation rate for STR3 of 0.45% per generation. Therefore, a change in STR3, be it a mutation or recombination, can occur approximately every 5555 years ([100/0.45]*25).

In addition to the STRs, two SNPs are known in exon 4 of *NCF1*, at nucleotide positions c.295 (rs62475423) and c.345 (rs79234723) (Fig. 2) [ref. 18,19]. SNP c.295 substitutes a G for an A, resulting in the change of an amino acid from glycine to serine. SNP c.345 substitutes a T for a C, but does not result in an amino acid change. The combinations of the alleles at these two SNPs (i.e., the SNP haplotypes) across the different groups can be used as a marker to investigate the possibility of a founder effect.

The aim of the present investigation was to determine 1) whether the mutation in exon 7 is present solely in the patients and their family members; 2) whether the haplotypes found in the 39 patients and their families differ from those found in other population groups (other Kavkazi Jews, regional population groups and Western Europeans), and 3) whether the number of repeats in STR1 and STR3 in the patients, family members and control groups can tell us something about the

Table 1
Population sizes.*

Code	Population	J/NJ	N	N comp	n haplos
CI	Controls Iran	NJ	50	15	30
CT	Controls Turkey	NJ	50	49	98
CWE	Controls Western Europe	NJ	52	50	100
CWG	Controls West Georgia	NJ	50	48	96
GJC	Georgian Jewish Controls	J	57	0	0
IJC	Iranian Jewish Controls	J	75	0	0
KC	Kavkazi Controls	J	226	108	216
KF	Kavkazi Family Members	J	27	26	52
KP	Kavkazi Patients	J	14	14	28
OF	Other Family Members	NJ	32	29	58
OP	Other Patients	NJ*	25	24	45
Total			658	363	723

J, Jewish; NJ, non-Jewish; N, number of individuals; N comp, number of completely analyzed individuals; n haplos, number of completely analyzed haplotypes.
*) With the exception of two Jewish, non-Kavkazi patients from one family in Ethiopia in this group.

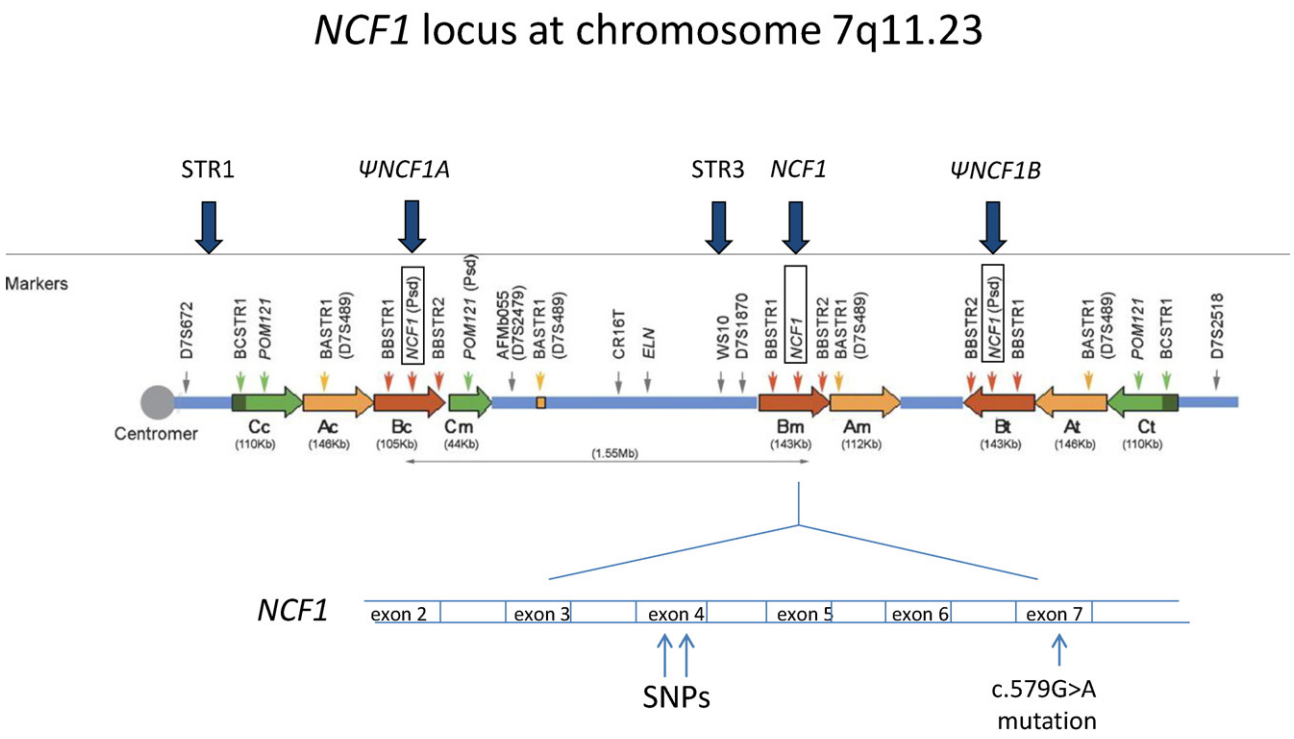

Fig. 2. Localization of *NCF1*, its two pseudogenes and STR1 and STR3 at chromosome 7q11.23. SNPs c.295 and c.345 are localized in exon 4 of *NCF1*, the mutation c.579G > A in exon 7 of *NCF1*.

homogeneity of the patient group and the date of introduction of the disease mutation in the Kavkazi patient group.

2. Materials and methods

2.1. Blood, cells, DNA

DNA from the patients and their relatives was sent by clinicians to our laboratory in Amsterdam for mutation analysis. Fourteen CGD patients were from Kavkazi Jewish ancestry in Israel, the others were from Turkey (eight), Italy (six), Jordan (three), Egypt (two), Ethiopia (two), Oman (one), Eritrea (one), Romania (one) and Malta (one). Seven control groups of Kavkazi Jews, Georgian Jews, Iranian Jews and non-Jewish individuals from West-Georgia, Turkey, Iran and Western Europe were used to compare the SNP and STR variations with those of the patients and their families. DNA from a group of Kavkazi Jews not related to the patients was obtained from Meir Medical Center, Kfar Saba, Israel, with the local Helsinki Ethics Committee approval, and from the Rambam Medical Center, Haifa, Israel, with IRB approval to identify carriers of the *NCF1* c.579G > A mutation. These Kavkazi Jews were identified by self-reported ancestry of three previous generations having lived in Dagestan or Azerbaijan. DNA from non-Jewish inhabitants of West-Georgia, Iran and Turkey was obtained from the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. Non-Jewish Western European EDTA blood samples were obtained from Sanquin, Amsterdam, The Netherlands, after appropriate informed consent had been obtained. All data were analyzed anonymously.

Genomic DNA was isolated from whole blood with the Gentra Puregene Kit (Qiagen, Hilden, Germany) or with QIAGEN QIAamp (Qiagen) according to the manufacturer's instructions.

2.2. Sequencing

Sequencing of exon 4 and exon 7 of *NCF1* (Fig. 2) was performed by Sanger sequencing according to [5]; reference for the sequence was retrieved from [20]. Primers were developed on *NCF1*-specific points in the introns of *NCF1* to amplify only *NCF1*- and not pseudo-gene-specific sequences (Suppl. Table 1).

2.3. STR length and repetition determination

From the hg19 assembly of the UCSC browser 4 Mb around *NCF1* (1 Mb 5′ of pseudogene *ψNCF1A* to 1 Mb 3′ of pseudogene *ψNCF1B* [Fig. 2]) was taken and used in Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.html). With this program three STRs were selected, each consisting of repeats of four base pairs. DNA samples were amplified by PCR with appropriate primers surrounding these STRs (Fig. 2) acquired from Invitrogen (Suppl. Table 1). As one STR did not yield good PCR products, we continued with the other two STRs. PCR product lengths were determined with the GeneScan method [21]. From the PCR product lengths, together with the hg19 reference sequence and primer design, the STR length was calculated. The resulting STR lengths were correlated to the number of tandem repeats. Coalescence times were calculated based on STR mutation rates [22,23].

2.4. Statistical analyses

Observed genotype counts in the control groups were compared to those expected assuming Hardy-Weinberg proportions with the Hardy-Weinberg Equilibrium (HWE) test for (multiallelic) markers from the gap package (version 1.1–12) [ref. 24]. Genotype counts were compared between populations with the Fisher exact test. Haplotype frequencies were estimated [25] and compared with the haplo.stats package (version 1.6.8) [ref. 26]. All analyses were carried out with R software (Version 3.0.3) [ref. 27]. Statistical significance was defined as a P-value <0.05.

3. Results

Table 1 shows the number of completely typed individuals per population investigated. For controls from Turkey (CT), controls from Western Europe (CWE), controls from West Georgia (CWG), family members

of Kavkazi Jewish patients (KF), Kavkazi Jewish patients (KP), family members of other patients (OF) and other patients (OP), nearly all individuals were genotyped at all five loci. In the group of Kavkazi Jewish controls (KC) 93 individuals were only typed for exon 7 of *NCF1*. Only completely genotyped individuals were included in the haplotype analysis. For Jewish controls from Georgia (GJC) and Jewish controls from Iran (IJC) no completely typed individuals were available (STR3 was not determined due to lack of sufficient DNA). All investigated patients had the c.579G > A mutation in *NCF1*. In the group of Kavkazi Jewish patients, one individual was heterozygous for the c.579G > A mutation and carried a splice site mutation in *NCF1* (intron2 + 1G > A) on the other allele. In the group of Other Patients, three individuals from two families were hemizygous for the c.579G > A mutation in *NCF1* and carried a complete deletion of *NCF1* on the other allele (copy number = 1), and two patients from one family were compound heterozygous for the combination of c.579G > A and c.124C > T (*p.Arg42Trp*) in *NCF1*. Of all the family members sampled, 6/58 tested negative for the c.579G > A substitution on both alleles. However, these family members were two aunts, two brothers, one sister and one father whose child's disease is caused by a compound heterozygous mutation. The other family members (52/58) were shown to be carriers of the mutation on one allele. Thus, all obligate carriers were found to be heterozygotes of the c.579G > A mutation. All patients lacked NADPH oxidase activity (DHR assay and/or NBT slide test) and did not show $p47^{phox}$ expression by their neutrophils (ref. 5 and not shown).

Table 2 shows the genotype counts and allele frequencies per population of the single nucleotide variants in *NCF1*. The genotype counts for the SNPs c.295 and c.345 in the control populations are not in Hardy-Weinberg equilibrium. Based on the observed minor allele frequencies, more heterozygotes and less homozygotes for the minor allele are expected, especially for c.345. It is obvious that the c.295G allele and the c.345 T allele are minor alleles in the control populations, but not in the patients (>90% G, >80% T) and their relatives (>50% G, >50% T). Even more striking is the total absence of the M (mutation) allele (c.579A) in the control populations, except for the Kavkazi Jewish Controls, where it was found in 2.2% of the alleles. In 27/28 families (96%) the c.579A > T mutation in *NCF1* is in one haplotype with c.295G and c.345 T (G_T_M). In three patients from one Turkish family, the G at the c.295 position was similar to the other patients, but at the c.345 location these three patients were homozygous for a C (G_C_M). The complete patient group did not display any SNP genotypes other than these two combinations.

Table 3 shows the total allele numbers and frequencies of the two STRs. For both STRs, the frequencies in the control populations were in Hardy-Weinberg equilibrium. Notable in these results is that 85.7% of the STR1 alleles observed in Kavkazi Jewish patients are 10 repeats of STR1 (24/28 alleles). Ten individuals were homozygous 10/10 and all individuals in this group carried at least one copy of the 10-repeat allele. Also in the family members of the Kavkazi Jewish patients (carriers of the *NCF1* c.579G > A mutation), the STR1 10-repeat was the most frequent allele found (59.5%, 32/54 alleles). In the Other, non-Jewish Patients (OP) the STR1 allele 10 was less frequently observed (32.0%). In total, three different STR1 alleles were found in Kavkazi Jewish patients and five in other patients. STR1 alleles 7, 8 and 15 were rarely observed in any population.

The number of STR3 repeats was determined in all groups except in the Georgian Jewish and Iranian Jewish groups, of which there was insufficient material for testing. The most frequently found STR3 allele in all populations was the 83 allele, in the Kavkazi Jewish patients even in 96% of all STR3 alleles. In these patients, only two different STR3 alleles were observed, in the other patients seven. The STR3 74, 86 and 87 alleles were only rarely observed in any population. Clearly, the group of Kavkazi Jewish patients is genetically much more homogeneous in the *NCF1* gene cluster than the group of other, non-Jewish patients. Also, all family members of the Kavkazi Jewish patients (carriers of the *NCF1* c.579G > A mutation) and all Kavkazi Jewish control individuals who were carrier of the mutation had an allele with 83 repeats for STR3.

In the Turkish family with three patients with the G_C exon 4 haplotype mentioned earlier, the father had an STR1 9/10 genotype and the mother an STR1 9/12 genotype (Fig. 3). Two of the three patients in this family were homozygous for the STR1 9 allele, indicating that this

Table 2
Genotype counts and allele frequencies per population for the variants in *NCF1*.

		CI (NJ)	CT (NJ)	CWE (NJ)	CWG (NJ)	GJC (J)	IJC (J)	KC (J)	KF (J)	KP (J)	OF (NJ)	OP (NJ)
c.295	Genotype	Count										
	A/A	29	35	41	37	36	55	85	4	0	2	0
	A/G	14	14	9	9	9	17	39	17	1	16	2
	G/G	5	0	2	3	6	2	8	6	13	12	20
	Allele	Frequency (%)										
	A	75.0	85.7	87.5	84.7	79.4	85.8	79.2	46.3	3.6	33.3	4.3
	G	25.0	14.3	12.5	15.3	20.6	14.2	20.8	53.7	96.4	66.7	95.7#
c.345	Genotype	Count										
	C/C	39	34	43	41	40	60	114	3	0	3	3
	C/T	8	14	7	4	8	13	14	18	1	14	2
	T/T	1	1	2	4	3	1	4	6	13	13	17
	Allele	Frequency (%)										
	C	89.6	83.7	89.4	87.8	86.3	89.9	91.7	44.4	3.6	33.3	17.0
	T	10.4	16.3	10.6	12.2	13.7	10.1	8.3	55.6	96.4	66.7	83.0*
Exon 7	Genotype	Count										
	W/W	50	50	52	50	34	66	215	4	0	2	0
	W/M	0	0	0	0	0	0	10	23	1^	30	2^
	M/M	0	0	0	0	0	0	0	0	13	0	20
	Allele	Frequency (%)										
	W	100.0	100.0	100.0	100.0	100.0	100.0	97.8	57.4	3.6	53.1	4.3
	M	0.0	0.0	0.0	0.0	0.0	0.0	2.2	42.6	96.4	46.9	95.7

#) three G alleles added or *) three T alleles added because three patients had only one *NCF1* allele present and are therefore not included in the genotyping.
W (wild type) indicates c.579G, M (mutation) indicates c.579A.
^) in the KP group is one patient and in the OP group are two patients with a combination of M on one allele and another mutation on the other allele.

Table 3
Allele frequencies per population for the two short tandem repeats (STR) (total allele number).

Variant		Population										
		CI (NJ)	CT (NJ)	CWE (NJ)	CWG (NJ)	GJC (J)	IJC (J)	KC (J)	KF (J)	KP (J)	OF (NJ)	OP (NJ)
		(100)	(100)	(104)	(100)	(28)	(28)	(262)	(54)	(28)	(62)	(50)
STR1	Allele	Frequency (%)										
	7	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	8	1.0	1.0	0.0	0.0	0.0	0.0	2.3	0.0	0.0	0.0	0.0
	9	19.0	21.0	26.9	19.0	10.7	28.6	22.5	7.4	10.7	16.1	20.0
	10	39.0	45.0	47.1	48.0	46.4	50.0	44.3	61.1	85.7	43.5	32.0
	11	6.0	0.0	0.0	2.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0
	12	10.0	10.0	9.6	5.0	14.3	3.6	13.4	5.6	0.0	9.7	8.0
	13	23.0	22.0	11.5	20.0	21.4	7.1	14.5	24.1	3.6	27.4	28.0
	14	2.0	0.0	4.8	6.0	3.6	10.7	2.7	1.9	0.0	3.2	12.0
	15	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	0.0	0.0	0.0
		CI	CT	CWE	CWG	GJC	IJC	KC	KF	KP	OF	OP
		(34)	(100)	(100)	(98)	(0)	(0)	(220)	(52)	(28)	(64)	(48)
STR3	Allele	Frequency (%)										
	73	11.8	18.0	17.0	8.2	–	–	13.2	9.6	0.0	17.2	18.8
	74	0.0	1.0	0.0	1.0	–	–	0.5	0.0	0.0	0.0	0.0
	77	0.0	0.0	0.0	3.1	–	–	3.2	0.0	0.0	0.0	0.0
	78	2.9	3.0	2.0	4.1	–	–	1.8	0.0	0.0	0.0	0.0
	79	0.0	0.0	2.0	1.0	–	–	1.8	0.0	0.0	9.4	8.3
	80	0.0	1.0	1.0	2.0	–	–	0.9	1.9	0.0	0.0	0.0
	81	2.9	3.0	3.0	1.0	–	–	0.0	1.9	0.0	0.0	4.2
	82	14.7	5.0	12.0	7.1	–	–	10.5	3.8	0.0	18.8	14.6
	83	47.1	43.0	27.0	38.8	–	–	23.6	57.7	96.4	29.7	31.3
	84	14.7	19.0	25.0	23.5	–	–	21.8	9.6	0.0	20.3	16.7
	85	5.9	7.0	11.0	10.2	–	–	21.4	13.5	3.6	3.1	6.3
	86	0.0	0.0	0.0	0.0	–	–	0.0	0.0	0.0	1.6	0.0
	87	0.0	0.0	0.0	0.0	–	–	1.4	1.9	0.0	0.0	0.0

was the *NCF1* c.579A-bearing haplotype. However, the third child had an STR1 9/10 genotype but nevertheless the exon 7 mutation in homozygous form. Thus, within this family, a cross-over has taken place between STR1 and *NCF1*.

We then estimated the different haplotype frequencies in the patients and their families and in the five control groups with sufficient information. Table 4 shows the results. We had information on all loci from 363 individuals. In the patient groups, three individuals carried a single haplotype (deletion of *NCF1* gene on the other allele), resulting in a total of 723 analyzed haplotypes. Of these, 10_83_A_C_W and 10_84_A_C_W were the most common "wild type" haplotypes (pooled frequency 11% and 8.2%). 10_83_G_T_M was the most frequently found mutant haplotype, only observed in KP (Jewish, J), KF (J), OP (Non-Jewish, NJ), OF (NJ) and KC (J) (in decreasing order). The ten most common haplotypes account for 26.5% (OP) to 85.6% (KP) of the haplotypes (estimated to be) present in the different populations, again illustrating the fact that the Kavkazi Jewish patients are genetically much more homogeneous than the Other Patients. It also illustrates that the Kavkazi Jewish patients form a separate group within the Kavkazi Jewish group as a whole. In accordance with this idea, the mutant haplotype 10_83_G_T_M was present in the Kavkazi Jewish controls for only 2%. When we analyzed whether the Kavkazi Jewish controls were genetically different from their geographic neighbors, we found that the estimated haplotype frequencies in KC (J) differed significantly from those in the combined CI (NJ) and CWG (NJ) populations ($p = 0.029$). At the individual locus level, allele frequencies for STR3 ($p = 0.019$) and exon 7 ($p = 0.034$) differed between KC and CI/CWG.

Finally, we selected the haplotypes containing the mutant *NCF1* c.579A allele. In total, 19 of such haplotypes had estimated frequencies larger than zero, and Table 5 shows the estimated frequencies of these haplotypes as percentage of the total number of mutant haplotypes. The 10_83_G_T_M haplotype was found most frequently in all four patient/family populations and also in the Kavkazi Jewish controls (KC, ten mutant haplotypes). In KP (J) only three mutant haplotypes were estimated to be present, i.e. 10_83_G_T_M, 13_83_G_T_M and 9_83_G_T_M. Thus, only STR1 differed between these haplotypes, possibly as a result of past recombination events. In contrast, in the other, non-Jewish, patients (OP) 11 different mutant haplotypes were observed, with 10_83_G_T_M, 13_82_G_T_M, 8_73_G_C_M, 14_84_G_T_M and 12_79_G_T_M all being present in more than 10% of all mutant haplotypes in these patients. Together, these five haplotypes made up more

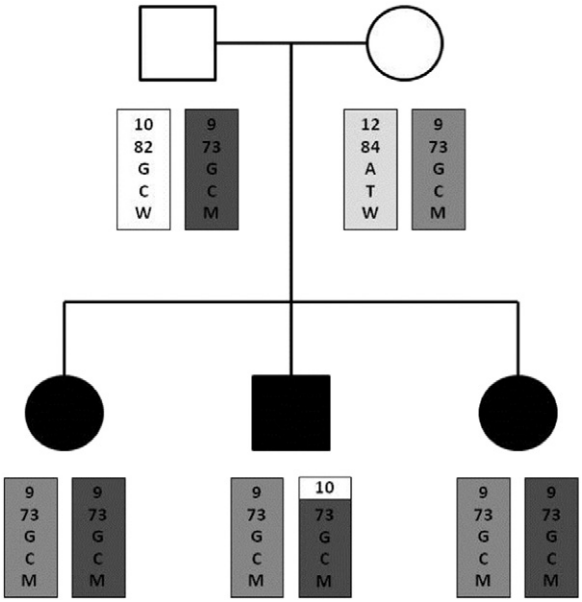

Fig. 3. Cross-over between STR1 and *NCF1* in a Turkish family.

Table 4
Estimated haplotype frequencies (%) of the ten most frequent haplotypes in the different populations (total haplotype number).

Haplotype	CI (NJ) (n = 30)	CT (NJ) (n = 98)	CWE (NJ) (n = 100)	CWG (NJ) (n = 96)	KC (J) (n = 216)	KF (J) (n = 52)	KP (J) (n = 28)	OF (NJ) (n = 58)	OP (NJ) (n = 45)
10_83_A_C_W	12.3	17.6	11.4	15.2	9.2	8.7	0.0	12.7	0.0
10_83_G_T_M *	0.0	0.0	0.0	0.0	2.0	36.2	85.6	13.3	22.1
10_84_A_C_W	6.3	10.2	11.9	9.5	8.3	2.0	0.0	8.5	0.0
10_85_A_C_W	2.1	2.7	4.6	5.5	11.0	5.1	0.0	0.7	0.0
9_83_A_C_W	0.2	7.9	5.6	6.8	2.5	0.1	0.0	2.0	0.0
9_84_A_C_W	2.7	2.8	4.0	4.5	4.4	0.0	0.0	0.0	0.0
9_85_A_C_W	1.9	1.6	3.0	2.3	3.8	3.4	0.0	1.0	4.4
10_73_A_C_W	2.1	2.7	5.9	2.6	2.5	2.7	0.0	0.3	0.0
13_83_A_C_W	4.4	6.2	1.4	2.5	2.3	3.8	0.0	0.0	0.0
9_73_A_C_W	2.4	2.3	5.3	0.0	3.5	0.0	0.0	0.0	0.0
Total	34.3%	54.0%	53.2%	48.8%	49.5%	62.1%	85.6%	38.5%	26.5%

*) The haplotype that carries the pathogenic mutation in exon7 of the *NCF1* gene.

than 70% of all mutant haplotypes estimated to be present in the OP. Also striking is the presence of the G_T_M SNP combination in 14 of the 19 mutant haplotypes. The other five SNP combinations observed were found in no more than 2.9% of the mutant haplotypes, except for the 9_73_G_C_M haplotype in 11.6% of the OP (due to five of these haplotypes in one family (Fig. 3).

4. Discussion

In this study we have investigated an isolated group of chronic granulomatous disease (CGD) patients of Kavkazi Jewish descent, and compared various genetic markers with their family members, with other CGD patients with the same mutation, and with control groups. The cause of CGD was common for all patients: a c.579G > A mutation in exon 7 of the *NCF1* gene that encodes the $p47^{phox}$ subunit of the NADPH oxidase enzyme complex. A total of 39 patients was examined, of whom 14 were Kavkazi Jews, and 25 were of non-Jewish descent (with the exception of two Jewish patients from one family) but all from the Middle East or South-East Europe. Investigation of the mutation in exon 7 showed that most patients have the mutation in homozygous form, three in hemizygous form (with deletion of the other allele), and three patients were compound heterozygous with another mutation in *NCF1*. In the control groups, we found the mutation in heterozygous form only, in 4.4% of the Kavkazi Jewish individuals (2.2% of their *NCF1* alleles).

Table 5
Estimated frequencies (%) of mutant haplotypes expressed as percentage of total number of mutant haplotypes (total mutant haplotype number).

Haplotype	KP (J) (n = 27)	OP (NJ) (n = 43)	KF (J) (n = 22)	OF (NJ) (n = 27)	KC (J) (n = 10)
10_83_G_T_M	88.8	23.1	85.7	28.5	48.6
13_82_G_T_M	0.0	16.3	0.0	23.0	0.0
9_73_G_C_M	0.0	11.6	0.0	0.2	0.0
9_84_G_T_M	0.0	4.7	0.0	10.9	0.0
13_83_G_T_M	3.7	9.5	10.6	2.5	4.7
14_84_G_T_M	0.0	9.3	0.0	7.4	0.0
12_79_G_T_M	0.0	9.3	0.0	3.7	0.0
13_73_G_T_M	0.0	6.8	1.2	1.0	0.0
9_83_G_T_M	7.4	0.0	0.7	0.0	14.1
13_79_G_T_M	0.0	0.0	0.0	6.9	0.0
14_83_G_T_M	0.0	4.7	0.0	0.0	0.8
10_73_G_T_M	0.0	0.2	0.5	4.3	0.0
10_85_G_T_M	0.1	2.3	1.1	3.7	0.7
9_86_G_T_M	0.0	0.0	0.0	3.7	0.0
9_82_G_C_M	0.0	0.0	0.0	2.9	0.0
12_77_G_C_M	0.0	0.0	0.0	0.0	11.1
14_80_G_T_M	0.0	0.0	0.0	0.0	9.0
10_73_G_C_M	0.0	2.5	0.0	1.1	0.0
12_83_A_T_M	0.0	0.0	0.3	0.0	11.1

4.1. Limitations

An important assumption of the expectation-maximization (EM) algorithm used to estimate the haplotype frequencies is that the genotypes are in Hardy-Weinberg equilibrium (HWE). In the present study, the observed genotypes at c.295 and c.345 are not in HWE. The observed excess of homozygosity, especially for the minor alleles, and deficit in heterozygosity, may be due to the lack of amplifying both *NCF1* genes: although the primers for exon 4 were developed on SNPs that are supposed to be different in *NCF1* and its two pseudogenes [2, 19], this is not always the case, leading to amplification of one instead of two *NCF1* sequences and thus to an overestimation of homozygous sequences.

4.2. Founder effect

In 92% of all patient haplotypes (96% of all mutated haplotypes in families with a c.579G > A CGD patient), the mutation in exon 7 was accompanied by the G_T combination of SNPc.295 and SNPc.345 in *NCF1*. This SNP combination was not found in homozygous form in any of the other groups that were investigated, except for family members of the patients and in one member of the Kavkazi Jewish control group. In general, we found a very strong linkage between the exon 7 mutation and the G_T combination in the same haplotype. However, the G_T combination without the mutation was found in the patients' families as well as in control groups. This is the first indication for a founder effect, probably resulting from the introduction of the c.579G > A mutation in an *NCF1* gene with the G_T SNP combination.

Further support for a founder effect in the Kavkazi group comes from the STR results. Although not uniform for all Kavkazi Jewish patients, 89% of the mutation-bearing haplotypes in this group have 10 repeats of the STR1 tandem. Analyzed per family this is even higher: 21/23 alleles = 91.3%. Two haplotypes were recorded with different numbers of repeats. This is a markedly higher consistency than found in any of the other groups, with 87% in the Kavkazi Jewish patients' family members being the next highest frequency for any other STR1 repeat. According to the combined recombination and mutation rate of 2.2% for STR1, the mutation in exon 7 would have been introduced less than 1135 years ago if all Kavkazi Jewish patients showed the same number of repeats. Even though this is not the case, the high consistency amongst the patients indicates that the mutation in exon 7 was introduced 1135 to 2270 years ago.

The results from STR3 provide more information about the introduction of the mutation in exon 7. The combined recombination and mutation rate for STR3 of 0.45% per generation over a distance of 0.25 Mb indicates that mutation or recombination occurs approximately every 5555 years. Such a change has not been noted in the Kavkazi Jewish patients, since all these patients show 83 tandem repeats in their mutated

haplotype. Also, all family members of the Kavkazi Jewish patients (carriers of the *NCF1* c.579G > A mutation) and all Kavkazi Jewish control individuals who were carriers of the mutation had a haplotype with 83 repeats for STR3. This observation is also consistent with the STR1 results. However, the wide distribution of the number of repeats in the other c.579G > A CGD patients indicates that the mutation in exon 7 has been present in individuals of non-Kavkazi Jewish descent for more than 5555 years or originated in several different haplotypes.

The accumulation of these data strongly supports the idea that a founder effect has been the cause of c.579G > A CGD in the Kavkazi Jewish population. One individual with the mutation in exon 7 of *NCF1*, the G_T SNP combination, an STR1 with 10 tandem repeats and an STR3 with 83 tandem repeats has probably introduced this genotype into the Kavkazi Jewish community between 1135 and 2300 years ago. The fact that other individuals were found with the same rare haplotype but without the disease-causing mutation in exon 7 of *NCF1* indicates that the mutation was originally introduced into this haplotype.

4.3. Exceptions

We found a few exceptions to this rule. One is a Turkish family with three children who had the G_C exon 4 haplotype and the mutation in exon 7 in homozygous form (Fig. 3). The mother was a carrier of the exon 7 mutation and heterozygous for G/A at position c.295 and for C/T at position c.345. This indicates the linkage between G_C and the mutation on the *NCF1* haplotype given to her children. The father was homozygous for the G_C haplotype but heterozygous for the exon 7 mutation. There are several options to explain the linkage between the exon-4 G_C haplotype and the exon-7 mutation in this family. One is that the mutation may have been introduced into the G_C haplotype, and another is a recombination of, for instance, a G_T_M haplotype with a G_C_W haplotype. The G_C_W haplotype is quite common in the Turkish population, but the exon 7 mutation was not found otherwise in the Turkish control group. Since this was a rare exception to the rule of the exon 7 mutation in the G_T exon 4 haplotype, we assume that the introduction of this mutation occurred rather recently.

Two other exceptions were found in the Kavkazi Jewish control group. One carrier of the exon 7 mutation had an A/G genotype at c.295 and C/C at c.345, indicating a link between the mutation and either A_C or G_C in exon 4. These SNP combinations are quite common in the Kavkazi Jewish control group. The other mutation carrier had an A/A & T/T SNP genotype, indicating linkage between A_T and the mutation. This SNP combination is found in about 10% of all Kavkazi Jewish control individuals. Table 5 shows that in the Kavkazi Jewish control group 19 different mutant haplotypes may exist: 4 with G_C_M, one with A_T_M and 14 with G_T_M.

Thus, the c.579G > A mutation is present in at least three different *NCF1* haplotypes. The most common of these is G_T, and this haplotype is the founder of all Kavkazi Jewish c.579G > A CGD patients. Most likely, the frequency of this mutation, which is found at a very low rate in the Middle East populations, was increased in the Kavkazi Jewish population due to a founder effect and population isolation. It started according to the statistical mutation rate analysis some 1200–2300 years ago, and - according to historical knowledge - during the establishment of the community in the 8th century C.E. It is unclear whether the mutation in the Kavkazi Jewish community is due to a *de novo* mutation in this community followed by a founder effect and isolation, or to the introduction of a variant present in the Middle East in low frequency into the Kavkazi Jewish community. The most abundant exon-4 – mutation combination is G_T_M. Presuming that this is the founder haplotype, the A_T_M combination (only found in the Kavkazi Jewish group) may have arisen from a recombination of G_T_M with A_T_W or by the introduction of the mutation in the A_T_W haplotype. Both G_T_M and A_T_W are present in the Kavkazi Jewish control group, at 3.2% and 3.3% of all haplotypes, respectively, so it cannot be decided which of these options is the most likely.

We also noticed that the two Jewish patients from one family in Ethiopia, although not belonging to the Kavkazi group, had one completely identical 10_83_G_T_M haplotype with that of the Kavkazi's. Whether this shared haplotype is due to shared ancestry remains to be established.

4.4. Expansion

One may wonder why such a high number of CGD patients (fourteen) and such a high percentage of carriers (4.4%) with the c.579G > A mutation is present in the Kavkazi Jewish group. This is perhaps explained by a bottleneck following expansion of the group to its present size of about 100,000 Kavkazi Jews in Israel (ref: Absorption Ministry of Israel). Indeed, there is a clear maternal founder event in Kavkazi Jews based on the evidence that 58% of the individuals have the J2b1 mitochondrial haplogroup [28]. It can be concluded that Kavkazi Jews constituted a population isolate with limited gene flow from other populations. Indeed, when we compared the haplotype frequencies in the Kavkazi Jewish control (KC) group with those in the neighboring Iranian, non-Jewish controls (CI) and West-Georgia non-Jewish controls (CWG) we found that this difference is significant ($p = 0.029$). For Kavkazi Jews the estimated time of expansion is 484 (133–1766) years ago. Therefore, it is quite feasible that the introduction of this mutation during a founder or bottleneck event followed by population expansion resulted in the observed frequency of the *NCF1* mutation introduced in this particular population.

It is unlikely that carrying this mutation confers a survival advantage for its bearers: the NADPH oxidase activity of these carriers is similar to that of non-carriers [29]. In contrast, patients with this mutation in homozygous or compound heterozygous form are prone to die at an early age from bacterial or fungal infections, before the age of reproduction. One would expect to find $0.045 \times 0.045 \times 0.25 = 0.0005$ or 5 : 10,000 patients in the Kavkazi group, or 50 in total in the whole group of about 100,000 Kavkazi Jews. We found 14, but know of an additional brother who died of CGD (according to his neutrophil dysfunctions) before DNA analysis was performed. Contemporary CGD patients are diagnosed earlier and with greater fidelity, and it is expected that more patients will be identified in the Kavkazi Jewish group. Therefore, we have launched a screening program in this group to identify carriers of the c.579G > A mutation in *NCF1*, and thus enable families to benefit from genetic counseling and early treatment.

5. Conclusion

A rare mutation in *NCF1* encoding $p47^{phox}$ of the leukocyte NADPH oxidase causes lack of superoxide generation, leads to chronic granulomatous disease and was recently (1200–2300 years ago) introduced into the Kavkazi Jewish population but present in surrounding populations already for more than 5000 years.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bcmd.2015.07.014.

Acknowledgements

We thank prof. Peter de Knijff (Leiden University, The Netherlands) for helpful advice and critical reading of the manuscript. The authors declare no conflict of interest.

References

[1] D. Roos, D.B. Kuhns, A. Maddalena, et al., Hematologically important mutations: The autosomal recessive forms of chronic granulomatous disease (second update), Blood Cells Mol. Dis. **44** (2010) 291–299.
[2] J. Roesler, J.T. Curnutte, J. Rae, et al., Recombination events between the *p47-phox* gene and its highly homologous pseudogenes are the main cause of autosomal recessive chronic granulomatous disease, Blood **95** (2000) 2150–2156.
[3] A. Hayrapetyan, P.C. Dencher, K. van Leeuwen, M. de Boer, D. Roos, Different unequal cross-over events between *NCF1* and its pseudogenes in autosomal

p47(phox)-deficient chronic granulomatous disease, Biochim. Biophys. Acta 2013 (**1832**) 1662–1672.
[4] D. Roos, S.M. Holland, T.W. Kuijpers, Chronic Granulomatous Disease, in: H.D. Ochs, C.I.E. Smith, J.M. Puck (Eds.), *Primary Immunodeficiency Diseases, a Molecular and Genetic Approach*. 3edn, Oxford University Press, New York, USA 2014, pp. 689–722.
[5] D. Roos, M. de Boer, M.Y. Köker, et al., Chronic granulomatous disease caused by mutations other than the common GT deletion in *NCF1*, the gene encoding the p47phox component of the phagocyte NADPH oxidase, Hum. Mutat. **27** (2006) 1218–1229.
[6] S. Al-Zadjali, S. Al-Tamemi, I. Elnour, et al., Clinical and molecular findings of Chronic Granulomatous Disease in Oman: Family studies, Clin. Genet. **87** (2015) 185–190.
[7] M. Altshuler, The Jews of the Eastern Caucasus: The History of the "Mountain Jews" from the Beginning of the Nineteenth Century, The Ben-Zvi Institute and the Institute of Contemporary Jewry, Hebrew University, Jerusalem, 1990. http://www.huji.ac.il/dataj/controller/ihoker/MOP-PUB_LINK?pub_id=027149&Save_t.
[8] N. Risch, H. Tang, H. Katzenstein, J. Ekstein, Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection, Am. J. Hum. Genet. **72** (2003) 812–822.
[9] R.I. Richards, G.R. Sutherland, Repeat offenders: Simple repeat sequences and complex genetic problems, Hum. Mutat. **8** (1996) 1–7.
[10] J.L. Weber, Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms, Genomics **7** (1990) 524–530.
[11] J.L. Weber, C. Wong, Mutation of human short tandem repeats, Hum. Mol. Genet. **2** (1993) 1123–1128.
[12] M.M. Albà, R. Guigó, Comparative analysis of amino acid repeats in rodents and humans, Genome Res. **14** (2004) 549–554.
[13] D. Metzar, C. Wills, Evidence for the adaptive evolution of mutation rates, Cell **101** (2000) 581–584.
[14] B. Brinkmann, M. Klintschar, F. Neuhuber, J. Hühne, B. Rolf, Mutation rate in human microsatellites: Influence of the structure and length of the tandem repeat, Am. J. Hum. Genet. **62** (1998) 1408–1414.
[15] H. Ellegren, Microsatellite mutations in the germline: implications for evolutionary inference, Trends Genet. **16** (2000) 551–558.
[16] D.B. Goldstein, C. Schlötterer, *Microsatellites*: *Evolution and applications*, Oxford University Press, Oxford, UK, 1999.
[17] G. Levinson, G.A. Gutman, Slipped strand mispriming: a major mechanism for DNA sequence evolution, Mol. Biol. Evol. **4** (1987) 203–211.
[18] A. Görlach, P.L. Lee, J. Roesler, et al., A p47-*phox* pseudogene carries the most common mutation causing p47-*phox*–deficient Chronic Granulomatous Disease, J. Clin. Invest. **100** (1997) 1907–1918.
[19] Database for Single Nucleotide Polymorphisms – Short Genetic Variations. NCBI. http://www.ncbi.nlm.nih.gov/projects/SNP/
[20] S.J. Chanock, J. Roesler, S. Zhan, et al., Genomic structure of the human p47-phox (*NCF1*) gene, Blood Cells Mol. Dis. **26** (2000) 37–46.
[21] J. Dekker, M. de Boer, D. Roos, Gene-scan method for the recognition of carriers and patients with p47(phox)-deficient autosomal recessive chronic granulomatous disease, Exp. Hematol. **29** (2001) 1319–1325.
[22] M.I. Jensen-Seaman, T.S. Furey, B.A. Payseur, et al., Comparative recombination rates in the rat, mouse, and human genomes, Genome Res. **14** (2004) 528–538.
[23] J.F. Chandler, Estimating per-locus mutation rates, J. Genet. Genomics Geneal. **2** (2006) 27–33.
[24] J.H. Zhao, gap: Genetic Analysis Package. R package version 1.1–12. http://cran.r-project.org/web/packages/gap/index.html 2014.
[25] R.M. Single, D. Meyer, J.A. Hollenbach, et al., Haplotype frequency estimation in patient populations: the effect of departures from Hardy-Weinberg proportions and collapsing over a locus in the HLA region, Genet. Epidemiol. **22** (2002) 186–195.
[26] J.P. Sinnwell, D.J. Schaid, Haplo.stats: Statistical analysis of haplotypes with traits and covariates when linkage phase is ambiguous. R package version 1.6.8. http://CRAN.R-project.org/package=haplo.stats 2013.
[27] R Core Team R, A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2013. http://www.R-project.org/.
[28] D.M. Behar, E. Metspalu, T. Kivisild, et al., Counting the founders: the matrilineal genetic ancestry of the Jewish diaspora, PLoS ONE **3** (2008) e2062.
[29] A.J. Verhoeven, M.L. van Schaik, D. Roos, R.S. Weening, Detection of carriers of the autosomal form of chronic granulomatous disease, Blood **71** (1988) 505–507.