# Testing life history theory in a contemporary African population

Meij, J.J.

**Summary**
Genetic association studies can only be successfully performed when the underlying genetic substructure of the population is known. Hidden population substructures have a major impact on the ability to detect significant associations, especially among traditional living African populations.
Here, we report the results of a detailed genetic survey among 205 males living in 93 compounds in a single village in the Garu-Tempane district of Ghana. All males were genetically screened for 15 autosomal STRs, 17 Y-chromosomal STRs, 27 biallelic Y-chromosomal markers defining Y-haplogroup E and sublineages thereof, and were sequenced for 365 bp of the mtDNA based HVR1 region 16024 – 16389.
We found that there are marked and significant Y-chromosomal genetic differences among the clans, but that the clans are not significantly different when analyzing autosomal STRs and mtDNA HVR1 genetic variation. This strongly suggests a highly reduced male mediated gene flow among the clans and a nearly fully random female mediated gene flow among the clans. This is confirmed by the very high mtDNA / Y-STR ration of Nע of 496 (332.33 / 0.669), and by analyzing the clan-specific Y-haplogroup distribution which was found to be highly significant non-random.
Except for members of the largest clan (Baakpang), most members of the other five clans display Y-STRs belonging to a single Y-haplogroup, and appear strongly clustered in the Y-STR network. Such a clear clustering was not observed for the HVR1 network.

## Introduction

The Bimoba are a relatively small agricultural tribe, living in the North-East corner of Ghana, Africa and parts of Togo. This is a remote and under developed area, with an average income far below the UN-poverty standard. In the Garu-Tempane district of Ghana, Leiden University Medical Center (LUMC) is actively involved in a number of medical genetic research projects among the Bimoba tribe. In addition to research into susceptibility for infections [1], the 'disposable soma theory' [2, 3] and parasitic load [4], also some basic modern healthcare is given to the local communities [5] and medical anthropological field work is carried out [6]. In addition to these research lines, a number of genetic association studies are currently in preparation. However, these can only be successfully performed, when the population genetic substructure is known. Hidden population substructure, reflecting differences in allele frequencies between cases and controls or reflecting differences in genetic variation between males and females, can have a major impact on the ability to detect significant association [7, 8]. Especially among traditional living African populations, sometimes displaying a marked difference in male and female mediated geneflow [9], such an insight is essential.

There are surprisingly few detailed African genetic studies on this topic, and those we know of only analyze the genetic differentiation among clearly distinct, different African populations widely dispersed throughout sub-Saharan Africa [10, 11]. To our knowledge, a detailed, tribal specific micro geographic study has never been performed. As such, this is surprising since there is some evidence that socio-cultural factors might play an important role in explaining differences in population demography, and at the same time influence genetic variation among closely related populations, thereby potentially confounding any genetic association study. Some insight into this problem might be extracted from the study performed by Destro-Bisol et.al. [12]. They analyzed in much detail the male-mediated and female mediated gene flow among sub-Saharan African food-producing-populations (FPP) and hunter-gatherer-populations (HGP). They concluded that socio-cultural factors, including polygyny and patrilocality, were responsible for much higher female mediated gene-flow (detected by reduced among population mtDNA genetic diversity) compared to a reduced male-mediated gene flow (as seen by a relatively increased among population Y-chromosome genetic variation) specifically among the FPP. However, they were only able to contrast and compare widely dispersed populations, and could not report on local (Tribe and or clan-specific) effects.

Here, we report the results of a detailed genetic survey among males living

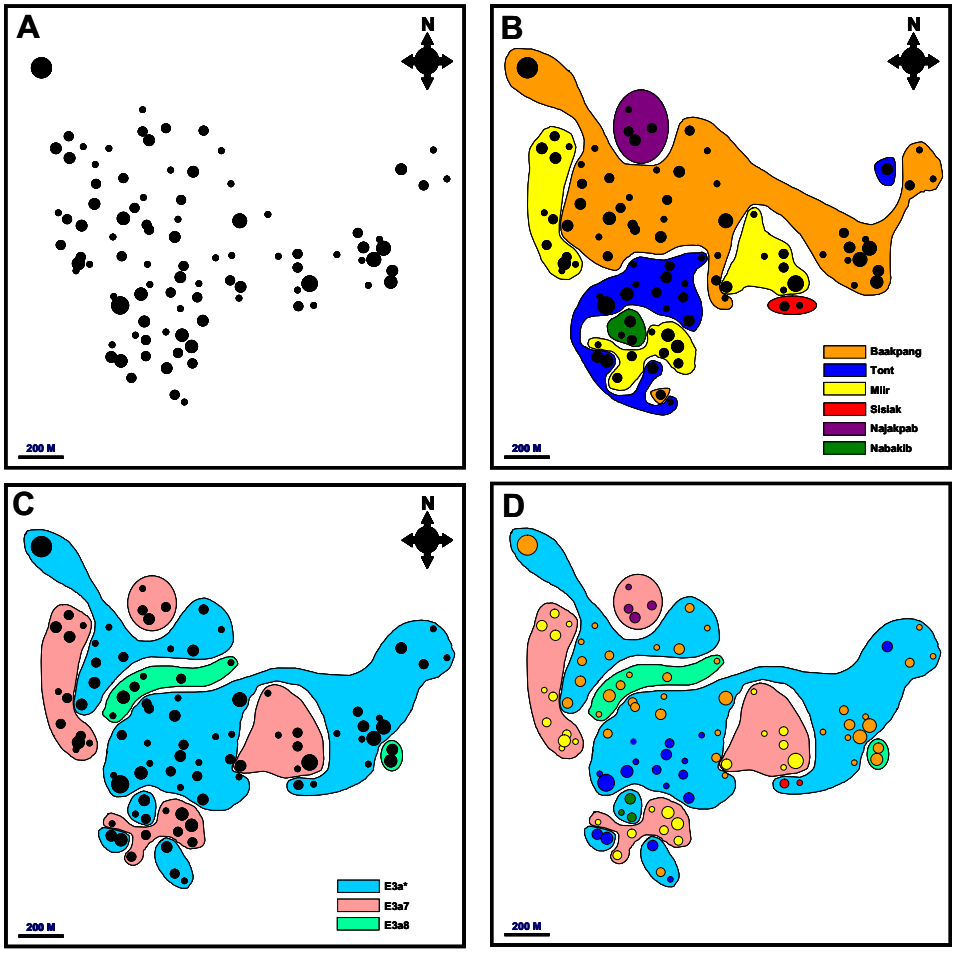


**Figure 1. (A)** Spatial distribution of the compounds analyzed in this study. Each dot represents a single compound. The relative diameter of each compound represents the number of males from each compound analyzed. The smallest dots represent a single male and the largest dot 8 males per compound. **(B)** Spatial distribution of the six clans. Each clan has a unique colour. This colour code is retained in panel (D) in this Figure, and in Figure 3. **(C)** Spatial distribution of the Y-SNP defined Y-haplogroups. Each Y-haplogroup has a unique colour. This colour-code is retained in panel D of this figure, in Figure 2 (page 62), and in Figure 3A (page 65). **(D)** The spatial distribution of Y-haplogroups superimposed over the spatial distribution of the clans. Each compound was given the colour-code of its clan members.

in a single village in the Garu-Tempane district. All males were screened for a set of autosomal short-tandem repeat (STR) loci, for Y-chromosomal STRs, and for a series of Y-chromosomal single nucleotide polymorphisms (SNPs). In addition we sequenced 379 bp of the mtDNA D-loop (the HVR1 region).

**Materials and methods**

*Research area and population* The study area is located in the Upper East Region of Ghana, between 0.226 W – 10.689 N and 0.81 W – 10.837 N [5]. Within this study area of approximately 360 $km^2$, there are about 17.500 inhabitants living in over 2300 individual compounds which are loosely clustered in 36 villages. About 66% of all individuals in this region belong to the Bimoba tribe. This tribe is spread throughout the upper west of Togo and the Upper East Region of Ghana. The second largest tribal group (29%) in this region is the Kusasi. The Upper East Region of Ghana in general, is an underdeveloped area (GNP $309 a year). Most of the inhabitants are traditional pastoralist. People live in family compounds which are essentially small farms. These farms produce at subsistence level and only small quantities of the crops are sold. The population forms a patrilocal and patrilineal structure: the women are accepted to their husbands' clan and the sons stay in or around their fathers' compound. It is custom not to marry inside your clan and polygamy is wide spread.

For the purpose of this study we concentrated on a single pure Bimoba village, called Farfar. Complete genetic data was obtained from 205 males living in 93 compounds (Figure 1A). In this village, members six different Bimoba clans are present:. Baakpang, Tont, Miir, Sisiak, Najabab and Naabakib. The Baakpang clan and the Tont clan form a single clangroup and so do the members of the Miir clan and the Sisiak clan [6].

*Geography* All villages and compounds were mapped with the Global Positioning System. Since there are no civil registries, all villages, compounds and individuals within the study area were registered and assigned a unique identification number. The name, sex, age, and tribe of each individual was registered.

*DNA isolation* DNA was isolated from buccal swap samples using the QIAamp mini kit (Qiagen, Inc., Chatsworth, CA), according to the manufacturers standard protocol.

*STR analysis* Powerplex®16 (for 15 autosomal STR-loci) and AmpFℓSTR®Yfiler™ PCR Amplification Kit (for 17 Y-chromosomal STR loci) PCR reactions were performed according to the manufacturers'

specifications. PCR products were analyzed using an ABI 3100 automated DNA sequencer and the Genemapper®ID software.

*mtDNA sequence analysis* We sequenced a fragment of 365 bp of mtDNA HVR1 (between positions 16024 and 16389) essentially as described (Parson 1998). Sequencing fragments were analyzed on an ABI 3100 automated DNA sequencer and the Seqscape® software. Sequences were manually aligned and edited using BioEdit vs. 7.0.5.2. Before analyses, the 10 bp long c-stretch (between 16084 and 16093) was removed from the aligned sequences.

*Y-SNP analysis* First, all males were screened for a core set of 19 Y-chromosomal SNPs, in order to allocate them to one of the Y-haplogroup E subgroups. These nineteen markers were drawn from the literature [14-17]. All of these markers define different groups within haplogroup E of the phylogenetic tree [18]. Loci M96, M33, M75, M2, M154, M191, M215, M35, M78, M224, M81 and M123 were taken from Underhill et al. [14]. Loci P2 and V6 were taken from Sanchez et al. [15]. Locus M281 was taken from Semino et al. [17]. Loci V12, V13, V22 and V32 were taken from Cruciani et al. [16]. For most markers we designed new primers, so that the lengths of the amplified genomic Y chromosome DNA fragments would range from 77 to 150 nucleotides to increase sensitivity and facilitating a single multiplex PCR and Snapshot detection. The sequence of each locus was obtained from GenBank® (http://www.ncbi.nlm.nih.gov) using BLAST. The primers for the genomic segments spanning one or more Y chromosome markers were designed with the Primer 3.0 program v. 0.2. All primers were selected to have theoretical melting temperatures near 60°C at a salt concentration of 180 mM and a purine:pyrimidine content close to 1:1, when possible. The lengths of the primers ranged between 20 and 27 nt. Primers with four or more bases at the 3' end complementary to another part of the primer were discarded or redesigned to avoid artefacts due to hairpin formation. Each primer pair was tested for primer-primer interactions, and the primer sequences were checked to avoid similarities with repetitive sequences or with other loci in the genome. Table 1 shows the sequences of the amplification primers selected.

HPLC purified primers for amplification were purchased from Biolegio. A primer stock was prepared by dissolving the lyophilized primers in purified water to a final DNA concentration of 10 pmol/µl. Each primer was tested in a singleplex PCR. Five ng template was amplified by PCR in a 12.5 µl reaction volume containing 1 x PCR buffer, 100 µM of each dNTP, 0.4 µM of each primer and 0.6 units of AmpliTaq Gold DNA polymerase at 94°C for

10 min followed by 35 cycles of 30 s at 94°C, 30 s at 60°C, 30 s at 72°C and a final extension for 5 min at 72°C.

**Table 1**. Y chromosome SNP's and Primer sequences for PCR amplification of 10 Y chromosome DNA fragments with SNPs.

| Locus | GenBank or dbSNPs accesion | PCR primers (5' → 3') Forward | Reverse | µM | Amplicon size (bp) |
|---|---|---|---|---|---|
| M96 | AC010889 | gccagccaagaatgaagaga | tgagctgtgatgtgtaacttgg | 0,1 | 150 |
| M33 | AC009977 | ccgtcataggctgagacaaga | ccccaagagagacaactgac | 0,15 | 150 |
| M75 | AC010889 | tgacttgtcaaaagccaaaaca | ttgaacagaggcatttgtga | 0,1 | 123 |
| P2 | AC010137 | gagaatcagctccagccatc | ttttggatcttcatgctggtt | 0,04 | 100 |
| M2 | Rs3893 | acggaaggagttctaaaattcagg | aaaatacagctccccctttatcct | 0,2 | 147 |
| M154 | AC010889 | aggctacaaattagtgcgaca | gaggcacagatacttaaaccattg | 0,06 | 77 |
| M191 | AC004474 | aaaaatggagtgtttatcagagctt | cccagacacaccaaaatatctc | 0,4 | 122 |
| M215 | AC006376 | tcaaactgttggtaaattttagagaaa | cagaagcatcagctggaaca | 0,25 | 97 |
| M35 | Rs1179188 | agggcatggtccctttctat | tccatgcagactttcggagt | 0,15 | 96 |
| M78 & M224 | AC010889 | tgcattactccgtatgttcgac | tggaagcttaccatcttttatga | 0,06 | 132 |
| M81 | Rs2032640 | gcactatcatactcagctacacatctc | ttgtttcttcttggtttgtgtga | 0,04 | 99 |
| M123 | AC010889 | gttgcccaggaatttgcat | cacagagcaagtgactctcaaag | 0,15 | 89 |
| V6 | AC012068 | gatggcacagtgttcgacag | cttctctccaaatgcctgct | 0,4 | 102 |
| M281 | AC010889 | agcaaagttgaggttgcaca | tgggcaacaccagaatctaa | 0,15 | 93 |
| V12 | AC012068 | ctgagttggattgttttaagttga | ttggtctctcttcatgtgctg | 0,15 | 150 |
| V13 | AC012068 | caacagtggaggacaaagca | aagaccagcctgaccaacat | 0,15 | 106 |
| V22 | AC012068 | tggcaatgcctcaacttaca | attccccaaggtttcagagg | 0,15 | 110 |
| V32 | AC012068 | gcaaatgttccatgaatggtg | ccagccagagaggcacttta | 0,4 | 111 |

In the multiplex reaction, the final setup of the PCR amplification included 5 ng DNA in a 12.5 µl reaction volume containing 1 x PCR buffer, 3.0 mM $MgCl_2$, 200 µM of each dNTP, 0.04-0.4 µM of each primer and 2.5 unit of AmpliTaq Gold DNA polymerase. All DNA amplifications were performed in a GeneAmp 9700 thermal cycler using the same program as was used for the monoplex reaction. The concentrations of the primers in the multiplex reaction were adjusted in order to obtain equal amount of each PCR product. The primer concentrations ranged from 0.04 to 0.4 µM. In order to eliminate the excess of primers and dNTPs, the PCR products were purified by adding 1.5 µl ExoSAP-IT-reagent (GE Healthcare, USA) to 2 µl of PCR product and incubated at 37°C for 30 min. The enzymes were inactivated at 80°C for 15 min.

Table 2 (next page) shows the genotyping primers designed for each SNP. Primers for detection of deletions and insertions were designed with the 3' base corresponding to the last base before the possible deletion or insertion. For each SNP system investigated in the present study, the following base would identify the polymorphism. The sequences of the primers were checked for the possibility of primer-dimer and hairpin formation and investigated in PCR without template ('self-extension reaction'). In order to distinguish between the sizes of the detection primers, the primers were synthesized with lengths between 22 and 64 nucleotides. The desired length of a primer was adjusted at the 5' end by addition of a piece of a 'neutral' se-

**Table 2.** Minisequencing primer sequences for typing of 19 Y-chromosome SNP markers

| Locus | Poly (dC) | Neutral sequence (5'→ 3') | Target specific sequence (5'→ 3') | Orientation | µM | Primer size | Mutation |
|---|---|---|---|---|---|---|---|
| M96 | None | None | GGAAAACAGGTCTCTCATAATA | Reverse | 0,06 | 22 | G/A |
| P2 | None | gacaa | AGGTGCCCCTAGGAGGAGAA | Forward | 0,06 | 25 | T/C |
| M81 | None | None | CTTGGTTTGTGTGAGTATACTCTATGAC | Reverse | 0,06 | 28 | G/A |
| M154 | None | acaa | GTTACATGGCCTATAATATTCAGTACA | Reverse | 0,1 | 31 | G/A |
| M191 | None | gaaagtctgacaa | AAAATATCTCATATTTTCAT | Reverse | 0,02 | 33 | A/G |
| M78 | None | aagtctgacaa | CTTATTTTGAAATATTTGGAAGGGC | Reverse | 0,03 | 36 | A/C |
| M215 | None | gtcgtgaaagtctgacaa | CAGCTGGAACAGTTAGAAAG | Reverse | 0,04 | 38 | C/T |
| M2 | None | cacgtcgtgaaagtctgacaa | TTCATTGTTAACAAAAGTCC | Reverse | 0,1 | 41 | G/A |
| M224 | None | cgtgaaagtctgacaa | AATTGATACACTTAACAAAGATACTTC | Forward | 0,03 | 43 | A/G |
| V6 | None | taggtgccacgtcgtgaaagtctgacaa | CCTGCTGCCGCATCTGCA | Reverse | 0,25 | 46 | T/C |
| M281 | None | gtgccacgtcgtgaaagtctgacaa | GCACAAACTCAGTATTATTAAAC | Forward | 0,1 | 48 | T/C |
| M33 | None | ccacgtcgtgaaagtctgacaa | CAGTTACAAAAGTATAATATGTCTGAGAT | Reverse | 0,15 | 51 | C/G |
| M75 | None | taggtgccacgtcgtgaaagtctgacaa | GAAAAGACAATTATCAAACCACATCC | Forward | 0,06 | 54 | C/T |
| M123 | None | taaactaggtgccacgtcgtgaaagtctgacaa | CATTTCTAGGTATTCAGGCGATG | Forward | 0,2 | 56 | T/G |
| M35 | None | actgactaaactaggtgccacgtcgtgaaagtctgacaa | TCGGAGTCTCTGCCTGTGTC | Reverse | 0,02 | 59 | G/A |
| V12 | None | tgacaa | TTGTGTAGATAATTCAAAGT | Reverse | 0,25 | 24 | C/T |
| V13 | None | cgtcgtgaaagtctgacaa | GCTCAAACTTCCCTTG | Reverse | 0,15 | 35 | A/G |
| V22 | 1 | aactgactaaactaggtgccacgtcgtgaaagtctgacaa | CCAAGGTTTCAGAGGTC | Reverse | 0,15 | 58 | C/G |
| V32 | 2 | aactgactaaactaggtgccacgtcgtgaaagtctgacaa | CACACATGTATATACACACC | Reverse | 0,25 | 64 | C/G |

**Table 3.** Y chromosome SNPs and primer sequences for PCR amplification of 8 Y chromo-some DNA fragments with SNPs

| Locus | GenBank or dbSNPs accesion | PCR primers (5' → 3') Forward | Reverse | µM | Amplicon size (bp) |
|---|---|---|---|---|---|
| U174 | AC010720 | tccctgcagtgaaatagttttg | tgcttcactgtctctgttcttca | 0.2 | 72 |
| U175 | AC010720 | ctggtcacactaaggcacca | gcctttgtcacaaggtgattt | 0.25 | 105 |
| U181 | AC010723 | tgtctttgttttggcaagga | gtttaccaggaacccccatc | 0.1 | 70 |
| M58 & M155 | AC009977 | ggccacccatttattgtctt | tgcaagactgacaggaaaaact | 0.2 | 150 |
| M116.2 | AC010889 | aaagagcttattagatgatagaaaaacat | tcaccaaaggaatgcacatc | 0.2 | 124 |
| M149 | AC010889 | ccaaaagttgaatggtaaaacaga | tcacctctttgtattggcttct | 0.2 | 144 |
| M10 | AC010137 | gctcatcattgcttagtttcca | tgtttctttcacttcaaattgcat | 0.2 | 123 |

**Table 4.** Minisequencing primer sequences for typing of 8 Y chromosome SNP markers

| Locus | Poly (dC) | Neutral sequence (5'→ 3') | Target specific sequence (5'→ 3') | Orientation | µM | Primer size | Mutation |
|---|---|---|---|---|---|---|---|
| U174 | None | None | GCATACCAGATTAACCCAT | Forward | 0.2 | 19 | G/A |
| U175 | 6 | aactgactaaactaggtgccacgtcgtgaaagtctgacaa | GACCAGGAGAAGTCAAGA | Reverse | 0.25 | 64 | C/T |
| U181 | None | None | TCTTTGTTTTGGCAAGGA | Forward | 0.15 | 18 | C/T |
| M58 | None | gacaa | AGGATTCCTCTCCATTCT | Reverse | 0.2 | 23 | T/C |
| M116.2 | None | aagtctgacaa | TTAAGATAATTAAAATTGACAG | Forward | 0.2 | 33 | T/A or T/G |
| M149 | None | ctaaactaggtgccacgtcgtgaaagtctgacaa | ACACTTAATAGAACACAAGC | Forward | 0.2 | 54 | G/A |
| M155 | None | gccacgtcgtgaaagtctgacaa | GAGAGGAATCCTCACCTATC | Forward | 0.2 | 43 | G/A |
| M10 | 4 | aactgactaaactaggtgccacgtcgtgaaagtctgacaa | GTTTATTCCCAATGATCTTA | Forward | 0.2 | 64 | T/C |

quence and, if necessary, a poly-C tail. The neutral sequence, 5'-AACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAA-3', is a random sequence that did not match with any human sequence in the NCBI non-redundant database [19]. Each detection primer is of unique length, in order to be able to distinguish SNP loci. This is of particular importance concerning SNP loci with the same nucleotide polymorphism, since SNP loci with different nucleotide polymorphisms would be detectable even with equal lengths.

Multiplex PCR minisequencing was performed in 5 µl reactions with 1 µl purified PCR product, 2.5 µl of SNAPshot™ reaction mix (Applied Biosystems, USA) and 0.02-0.25 µM of the primers (Table 2). The minisequencing reaction was performed in a GeneAmp 9700 thermal cycler using the following programme: pre-denaturation at 96°C for 2 min, followed by 25 cycles for 10 s at 96°C, 5 s at 50°C and 30 s at 60°C. A positive control (a known DNA profile of a researcher) and a negative control (sterile water) were performed for each batch of 93 samples. The homogeneity of each primer was checked in singleplex minisequencing. The occurrence of extra peaks indicated heterogeneity of the minisequencing primer. After the minisequencing reaction, 1.25 µl SAP-reagent (Shrimp Alkaline Phosphatase reagent, GE Healthcare, USA) was added to each sample and the batch was incubated at 37°C for 1 hour in order to remove the 5' phosphoryl groups of the unincorporated [F]ddNTPs. SAP was inactivated by incubating at 75°C for 15 minutes. Two µl of the purified minisequencing PCR product was analyzed on an ABI 3100 automated DNA sequencer with a 36 cm capillary array, POP-4 polymer and 22 secons at 1000 V injections. GeneScan-120 LIZ™ was used as internal size standard. The data were analyzed using GeneMapper Analysis software v. 2.0 (Applied Biosystems, USA). After background subtraction and color separation, peaks were sorted into bins according to sizes by comparison to the internal size standard.

After completing the multiplex reaction described above and the Y-STR analysis to be described later, it was necessary to further subdivide haplogroup E3a*. Markers U174, U175 and U181 were drawn from recent literature [20]. The complementary primers designed are shown in table 3. These markers were analyzed in a monoplex reaction. The PCR conditions were the same as they were in the singleplex PCR reaction for the markers of the multiplex described above. The purifying of the PCR products, the minisequencing reaction (primers shown in table 4) and the capillary electrophoresis were done the same way as for the multiplex reaction described above. A few samples of each different group of haplotypes were

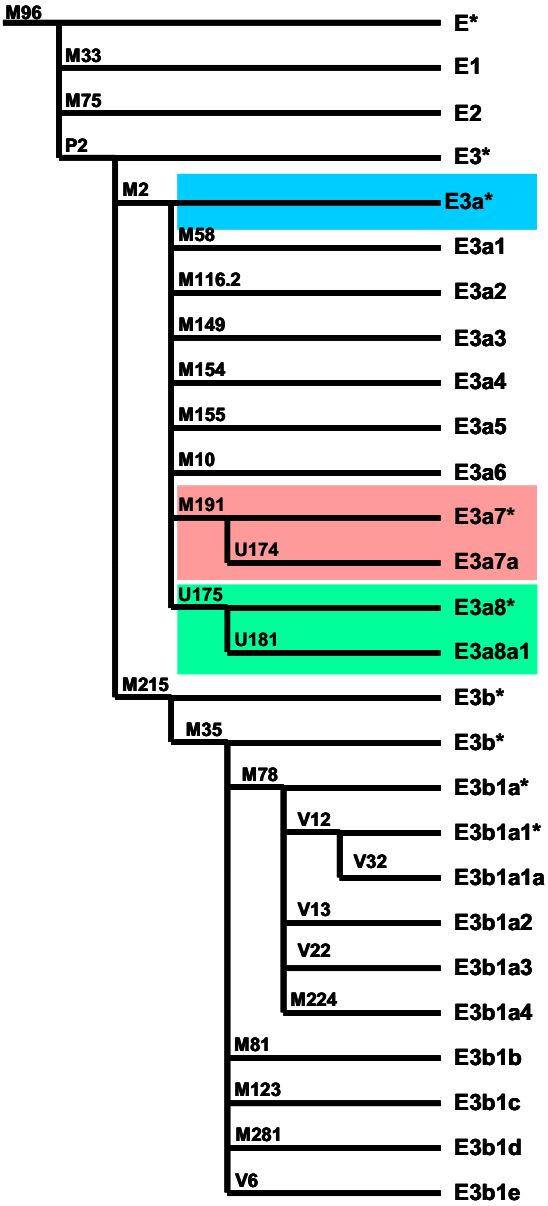


**Figure 2.** The phylogenetic tree of Y-chromosomal haplogroups. The phylogeny is based on Jabling et al (2003) [18] with some new additions. To the left of the tree the codes of all binary markers used to define each branch is given (i.e. M96, or V6). To the right of each branch, the uniform Y-haplogroup code is given. The only three different haplo-groups found among the Bimoba analyzed in study are indicated with the same colour-code as in Figure 1 and Figure 3A (see page 56).

tested for the markers M58, M116.2, M149, M155 and M10 [14]. The phylogenetic tree of all markers typed in this study is shown in Figure 2.

*Statistical analyses* We used Arlequin vs. 3.11 [21, 22] to estimate Fst and gene diversities among the different groups and for each different genetic system tested. We also estimated the parameter Nע (which incorporates effective population size, migration and mutation) by using the simple formula Nע = (1/Fst) – 1, according to the island model of migration for haploid systems [12,23]. Assuming that the effect of mutation rate is negligible, the different Nע estimates can be assumed to be the result of differences in migration rate and or effective populations size. When applied to Y-chromosome and mtDNA genetic data, different Nע estimates reflect differences in male and female mediated gene flow among the different clans. We used Network vs 4.2.0.1.[24, 25] to draw median joining networks based on the combined Y-STR and Y-SNP information and HVR1 sequence variation. For both genetic systems, a variable weight was given to different variable loci/sites. In order to estimate these different weights, we first drew a network giving all positions and equal weight and used the statistics option on the fully drawn network to obtain an estimate of the rate of homoplasy. Based on these estimates, highly homoplasic positions were down weighed accordingly.

**Results**

In this study we were able to analyze the DNA of 205 males from the Bimoba tribe, from the north-east Ghanaian village Farfar. These males belong to 6 different clans: Baakpang (n=90), Tont (n=43), Miir (n=55), Sisiak (n=3), Najakpab (n=8), and Nabakib (n=6). Males were living in 93 different compounds scattered over an area of approximately 4 $km^2$ (figure 1A). All males were genetically screened for 15 autosomal STRs, 17 Y-chromosomal STRs, 27 biallelic Y-chromosomal markers defining Y-haplogroup E and sublineages thereof, and were sequenced for 365 bp of the mtDNA based HVR1 region 16024 – 16389.

We first studied the spatial distribution of the various clans within this village (Figure 1b, page 56). No compounds were found to harbor members of multiple clans. Also, there is a clear non random distribution of clan-specific compounds within this village. Different clans cluster strongly together, and only when there is land-shortage, new clan compounds are settled outside the core clan-area.

Next we studied the genetic variation within and among the different clans. For this we estimated within clan and among clan genetic variation by means of Fst, AMOVA components and gene diversity across all loci. The results of these analyses are shown in Tables 5 and 6 (next page). It is very obvious that

**Table 5.** Fst and gene diversity estimates for each clan based on Y-STRs, autosomal STRs, and HVR1 sequences.

| | | Fst | | | Gene diversity | | |
|---|---|---|---|---|---|---|---|
| **Clan** | **Nr-ind.** | **Y-STR** | **Au-STR** | **HVR1** | **Y-STR** | **Au-STR** | **HVR1** |
| **Baakpang** | 90 | 0.593 | 0.012 | 0.004 | 0.196 | 0.783 | 0.017 |
| **Tont** | 43 | 0.607 | 0.012 | -0.001 | 0.058 | 0.776 | 0.020 |
| **Miir** | 55 | 0.599 | 0.012 | 0.001 | 0.142 | 0.784 | 0.019 |
| **Sisiak** | 3 | 0.610 | 0.016 | 0.006 | 0.044 | 0.716 | 0.024 |
| **Najakpab** | 8 | 0.610 | 0.013 | 0.008 | 0.033 | 0.773 | 0.018 |
| **Nabakib** | 6 | 0.613 | 0.014 | 0.013 | 0.000 | 0.739 | 0.015 |

**Table 6.** Analyis of molecular variance (AMOVA) and migration (Nγ).

| | Within clans | Among clans | | |
|---|---|---|---|---|
| **Markers** | **Variance (%)** | **Variance (%)** | **Fst** | **Nγ** |
| **Y-STRs** | 40.9 | 59.9 | 0.599 | 0.669 |
| **Au-STRs** | 98.8 | 1.19 | 0.012 | |
| **HVR1** | 99.72 | 0.28 | 0.003 | 332.33 |

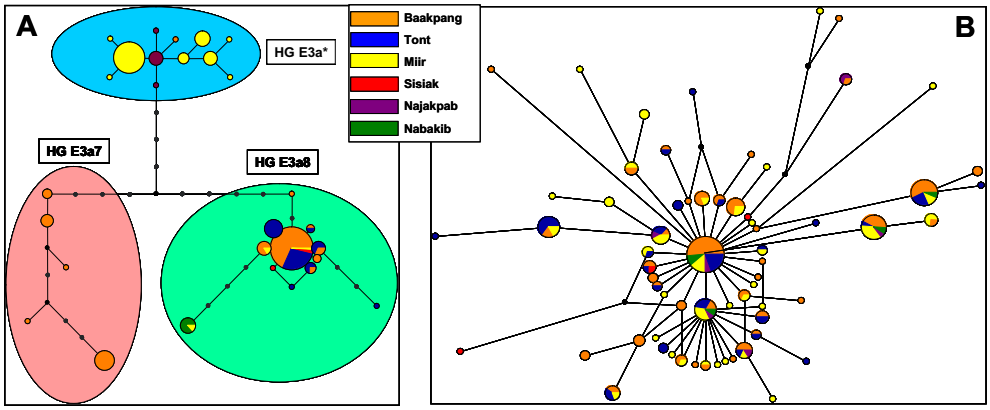


**Figure 3**. **(A)** Median joining networks connecting the Y-STR defined haplotypes, and **(B)** mtDNA HVR1 sequence haplotypes. The diameter of each pie is a relative indication of the frequency of that pie. The smallest pies represent a single individual. Each pie-segment colour corresponds to the clan colour code used in Figure 1 (page 56). Segments of the Y-haplotype network containing specific Y-haplogroups are indicated by the colour code for that Y-haplogroup as in Figure 1 and Figure 2 (page 62).

there are marked and significant Y-chromosomal genetic differences among the clans, but that the clans are not significantly differentiated when analyzing auto-somal STRs and mDNA HVR1 genetic variation. This is expressed by the very high among population genetic variance (59.9%) observed for Y-STRs, compared to the very low estimates for autosomal STRs (1.19%) and HVR1 (0.28%) (Table 6). This strongly suggests a highly reduced – if at all - male mediated gene flow among the clans and a nearly fully random female mediated gene flow among the clans. This is confirmed by the very high mtDNA / Y-STR ration of Nע of 496 (332.33 / 0.669) (Table 6). This marked gender specific difference in geneflow among the clans is also reflected in the distribution of the various Y-STR haplotypes and HVR1 sequence haplotypes across the clans (Figure 3). This is also confirmed by analyzing the clan-specific Y-haplogroup distribution (Table 7). This distribution is significantly non-random ($p<0.001$, Monte-Carlo Fisher exact test). Except for members of the largest clan (Baakpang), most members of the other five clans display Y-STRs belonging to a single Y-haplogroup, and appear strongly clustered in the Y-STR network. Such a clear clustering is not observed in the HVR1 network. The Y-related clan-specific correspondence is also very obvious when plotting and combining the Y-haplogoup distribution across the clan-specific compound distribution (Figure 1C and 1D, page 56).

**Table 7.** distribution of Y-haplogroups among Clans in Farfar.

| | Y-haplogoup (n and (% of total males)) | | |
|---|---|---|---|
| **Clan** | **E3a*** | **E3a7** | **E3a8** |
| Baakpang | 1 (0.5) | 22 (10.7) | 67 (32.7) |
| Tont | 0 (0) | 0 (0) | 43 (21.0) |
| Miir | 52 (25.4 | 0 (0) | 3 (1.5) |
| Sisiak | 0 (0) | 0 (0) | 3 (1.5) |
| Najakpab | 8 (3.9) | 0 (0) | 0 (0) |
| Nabakib | 0 (0) | 0 (0) | 6 (2.9) |

**Discussion**

It is known for quite some time, that there are significant gender specific differences in a number of demographic processes among various African populations [9, 10, 12]. Such differences were found to depend strongly on the type of population. Among traditional hunter-gatherer populations (HGP), such as central African pygmies, the female mediated gene flow (detected by means of mtDNA variation) is substantially reduced compared to male mediated gene flow (as detected by y-chromosomal genetic variation patterns) [12]. The reverse is generally observed among traditional pastoral farming communities. This is usually attributed to the combined influence of patrilocality and polyginy which appears to be the dominant type of society structure among African farming groups [12]. All studies so far have analyzed genetic differences among widely dispersed sub-Saharan populations, and did not consider differences among clans within the same tribe and/or village. One could argue that significant general trends might only be observed when comparing dispersed populations, but such trends might not be relevant when studying populations at the micro-geographic scale. Here we report the first results of such a micro-geographic study among 205 male Bimobas from six different clans living in the Upper East Ghanaian village of Farfar. Based on a detailed anthropological study [6], it was reported that among the Bimoba, clan and clangroup structure still plays a vital role in many cultural and demographic aspects of daily life. For instance, land is owned by the clan and not by the individual, and the mutual support system is restricted to the clan. Religious life and rite de passage are clan-based and varies between the clans.

We were able to compare the differences in male mediated gene flow and female mediated gene flow among the six clans in Farfar by analyzing Y-chromosomal and mtDNA genetic variation patterns in some detail. We found a markedly skewed male population substructure due to an almost complete lack of male geneflow among clans and a virtually random female geneflow among clans. It seems very likely that this peculiar genetic substructure is the immediate result of the strong clan and clan-group dominated social structure among the Bimoba's. It was remarkable to see such a strong Y-haplogroup difference among the different Bimoba clans. This could perhaps indicate different regions of origins for some of clans present in this single village. There is very little known about the origins of the Bimoba. The Moba, closely related to the Bimoba, migrated from Sudan to the west of Africa and it is clear that some clans of the Bimoba (the Naniik, Kpikpira and Nabakib clans) were sub-groups of the Moba. There is no clear indication when the Moba or Bimoba actually came to the west, but oral history claims that they did so in the aftermath of fights at the end of the

Shilluk reign, 1500 AD [6]. They all settled along the route from Sudan to Ghana. The Bimoba settled at the end of the line and claim that they have migrated from the Sudan separately. They seem to originate from nomad traders. Some other clans (Tambiouk, Maab, Bakpang and Tont) came, according to oral history, from the area that presently is known as south Togo and from the South. (Ashanti en Dagomba land).

Since we were only able to study a single village so far, it is too premature to speculate on the general consequences of our findings. For this we need similar data from more villages, different tribes in the same region, and detailed pedigree information in order to verify paternities and maternities. Such studies have recently initiated and will hopefully provide a firm support for our unexpected results.

**Acknowledegement**
We thank the members of the Garu-Tempane district, UER Ghana and the Ghanaian and Dutch field staff for their enthusiasm, their kind participation and hospitality during our fieldwork.

## References


1. Meij JJ, de Craen AJM, Sichiman HED, May L, Amankwa J, et al (2007) A contemporary study into ancient selection for Caspase 12 gene variants in rural Africa. *Chapter 7 of this thesis.*

2. Meij JJ, van Bodegom D, Ziem JB, Amankwa J, Polderman AM, et al (2007) Quantity trades off with quality of human offspring unexplained by differences in environmental conditions. *Chapter 5 of this thesis*.

3. Westendorp RG, Kirkwood TB (1998) Human longevity at the cost of reproductive success. Nature 396: 743-746.

4. Gasser RB, de Gruijter JM, Polderman AM (2006) Insights into the epidemiology and genetic make-up of Oesophagostomum bifurcum from human and non-human primates using molecular tools. Parasitology 132: 453-460.

5. Ziem JB, Olsen A, Magnussen P, Horton J, Agongo E, et al (2005) Distribution and clustering of Oesophagostum bifurcum and hookworm infections in Northern Ghana. Parasitology 132: 525 – 534.

6. Meij JJ, van Bodegom D, Baya Laar D (2007) The Bimoba. The people of Yennu. *Chapter 2 of this thesis*.

7. Freedman, M.L. et al. (2004) Assessing the impact of population stratification on genetic association studies. Nat Genet, 36: 388-393.

8. Marchini, J. et al. (2004) The effect of human population structure on large genetic association studies. Nat Genet, 36: 512-517.

9. Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, et al (2005) Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. Eur J Hum Genet 13:867-876.

10. Tishkoff SA, Williams SM (2002) Genetic analysis of African populations: Human evolution and complex disease. Nat. Rev. Genet. 3: 611 –621.

11. Adeyemo A, Chen G, Chen Y, Rotimi C (2005) Genetic structure in four West African population groups. BMC Genet 6:38.

12. Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, et al (2004) Variation of female and male lineages in sub-Saharan poulations: the importance of sociocultural factors. Mol Biol Evol 21: 1673 – 1682.

13. Parson W, Parsons T, Scheithauer R, Holland MM (1998) Population data

for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: Application of mtDNA sequence analysis to a forensic case. Int. J. Leg. Med. 111: 124 – 132.

14. Underhill PA, Passarino G, Lin AA, Shen P, Mirazón Lahr M, et al (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. 65: 43-62.

15. Sanchez JJ, Hallenberg C, Børsting C, Hernandez A, Morling N (2005) High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. Eur. J. Hum. Genet. 13: 856-866.

16. Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R (2006) Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. Hum. Mut. 27: 831-832.

17. Semino O, Silvana Santachiara-Benerecetti A, Falaschi F, Luca Cavalli-Sforza L, Underhill PA (2002) Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am. J. Hum. Genet. 70: 265-268.

18. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat. Rev. Genet. 4: 598-612.

19. Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN et al (2000) Large-scale discovery and genotyping of single nucleotide polymorphisms in the mouse. Nat. Genet. 24: 381-386.

20. Sims LM, Garvey D, Ballantyne J (2007) Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. Hum Mutat. 28: 97.

21. Arlequin version 2.000: [http://anthropologie.unige.ch/arlequin] a software for population genetics data analysis. Schneider S, Roessli D, Excoffier L. 2000.

22. Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 1992, 131:479-491.

23. Cavalli-Sforza LL, Bodmer W (1971) The genetics of human populations. Freeman, San Fransisco, California.

24. Network 4.1.1.2 - Fluxus Techn. Ltd. 2004. [www.fluxusengineering.com].

25. Bandelt H-J, Forster P, Röhl A: Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 1999, 16:37-48.