Insights into the genetic epidemiology of SBMA: prevalence estimation and multiple founder haplotypes in the Veneto Italian region.

DR. CINZIA BERTOLIN (Orcid ID : 0000-0002-4034-8170)

DR. GIORGIA QUERIN (Orcid ID : 0000-0001-6938-6368)

Article type : Original Article

Insights into the genetic epidemiology of SBMA: prevalence estimation and multiple founder haplotypes in the Veneto Italian region

*Bertolin C[1], *Querin G[2], Martinelli I[1], Pennuto M[3], Pegoraro E[1], Sorarù G[1]

1. Department of Neurosciences, University of Padova, Padova, Italy.

2. Laboratoire d'Imagerie Biomédicale, Sorbonne University, CNRS, INSERM, Paris, France.

3. Department of Biomedical Sciences, University of Padova, Padova, Italy.

* These authors contributed equally to this work

Corresponding author:

Gianni Sorarù

Neuromuscular Centre

Department of Neurosciences,

University of Padova, 35128 Padova, Italy

Telephone: +39 (0)49 8216394

Fax: +39 (0)49 8751770

E-mail: gianni.soraru@unipd.it

Keywords: SBMA, epidemiology, AR, genetic haplotypes, founder effect

Running Title: Genetic epidemiology of SBMA

Abstract

Background And Purpose: Literature data on SBMA epidemiology are limited and restricted to specific populations. Aim of our study was to accurately collect information about SBMA patients living in the Veneto region in Italy to compute reliable epidemiological data. AR lineages were genotyped to evaluate presence of founder effect.

Methods: We carried out a prevalence survey considering all SBMA patients diagnosed in the Italian Veneto region on January 31st, 2018. We evaluated the presence of different haplotypes genotyping 15 polymorphic markers (SNPs and STRs) around the AR gene.

Results: Based on 68 patients, the punctual prevalence of the disease on January 31st, 2018 was 2.58/100.000 (IC95% 1.65-3.35) in the male population. We identified 5 different haplotypes, confirming the existence of multiple founder effects. We also observed that, within the same haplotype, patients had a similar CAG repeat number ($p < 0.001$).

Conclusions: We calculated a reliable estimation of SBMA prevalence in the Italian Veneto regions which does not seem to be affected by strong founder effect. Moreover, our data suggest that length of CAG expansion could be preserved in patients harbouring the same haplotype.

Background

Spinal and bulbar muscular atrophy (SBMA), also known as Kennedy's disease, is a genetically determined, adult onset, slowly progressive neuromuscular disease. It is characterized by weakness, fasciculations and atrophy due to degeneration of lower motor neurons [1], but a concomitant primary myopathy has been also described [2, 3], along with a wide spectrum of systemic disorders [4, 5, 6, 7]. SBMA is an X-linked recessive inherited disease caused by a CAG repeat expansion, encoding a poly-glutamine (polyQ) tract, in the first exon of the androgen receptor (*AR*) gene. Thirty-eight or more CAG repeats are considered pathogenic. PolyQ-AR toxicity is hormone-dependent and symptoms of the disease appear only in adult male subjects. Clinical presentation of the disease may be extremely heterogeneous [8], with some patients presenting only sub-clinical or unspecific symptoms such as hand tremor and muscle cramps [9]. For this reason, obtaining exhaustive epidemiological information is difficult and an underestimation of the prevalence of the disease is probable. At the moment, only two studies describing the epidemiology of SBMA are available and both consider small patient cohorts. One of them focused on the Vasa District in Finland, considering 13 affected subjects from 10 different families and estimating the prevalence of the disease at 15/100.000 male inhabitants [10]. The second study involved 11 patients from a little Italian area (Reggio Emilia district) [11] and estimated the prevalence of the disease at 3.3/100.000 inhabitants. The variability in the local prevalence of the disease could be partially justified by the presence of a founder effect, which has been already described in Scandinavian, Japanese and mixed patients' populations [12-14].

Aim of our study was to accurately identify and collect information about SBMA patients living in the Veneto region in Italy in order to compute reliable epidemiological data. AR lineages were also genotyped to gain insight into generation of pathological states.

Methods

Geographical area of the survey

The Veneto region is situated in the North-East of Italy and extends over 18407,42 km^2. At January 31st, 2018, the resident population was estimated at 4.906.400 inhabitants of which 2.393.923 were male (data from the Italian National Institute of Statistic, ISTAT). The region is subdivided in 7 districts: Belluno (204.652 inhabitants of which 98.960 are men), Padova (936.274 inhabitants of which 455.386 men), Rovigo (238.588 inhabitants, 115.637 men), Treviso (885.972 inhabitants, 434.507 men), Venezia (854.275 inhabitants, 413.281 men), Verona (921.557 inhabitants, 450.834 men) and Vicenza (865.082 inhabitants, 425.318 men).

Study population

The study population included all the patients genetically diagnosed with SBMA in the Veneto Region (Italy) on January 31st 2018. The patients expressed informed written consent to the study, which was previously approved by the Ethical Committee of the Azienda Ospedaliera of Padova (Italy).

Cases were ascertained from the "Veneto Regional Registry for Rare Diseases" which is a prospective epidemiological registry involving all the neuromuscular centers of the region. The register was established in 2001 and is still ongoing [15].

Moreover, systematic genetic test for SBMA was performed during the years 2001-2017 in patients referred to our Neuromuscular Centre at the University of Padova (Italy) to further

investigate various neurological symptoms, such as muscle weakness, tremor, paraesthesia, and in subjects presenting neurophysiological findings compatible with lower motor neuron degeneration.

We collected family trees for at least 3 generations by direct interview of patients, going back until the beginning of the 20th century (range 1877-1929). The pedigrees allowed also the verification of the geographical origin of the families, as represented in the map in supplementary figure 1.

the neurology depart-

ments, (b) the motor neurone disease associations

(MNDA), (c) acute hospital trust coding system lists,

(d) the regional pharmacy unit in the Royal Victoria

Hospital, (e) general practitioners (GPs) and (f) neu-

rophysiology departments

Only patients living in the Veneto Region were considered for the epidemiological analysis, regardless from their place of birth.

Demographic and clinical history data including age and site of symptoms onset, age at diagnosis of the disease and age of death as well as information about eventual death's causes were collected.

Genetic analysis

Genomic DNA was extracted from peripheral blood

using standard protocols. PolyQ (CAG repeats) and

polyG (GGN repeats) alleles were amplified by poly-

merase chain reaction as previously described [5].

Repeat fragment sizing was performed on an ABI

Prism 3700 DNA Sequencer (Applied Biosystems,

Foster City, CA, USA). The specific length of both

the CAG and GGN repeats was further verified via

Sanger sequencing.

Genomic DNA was extracted from peripheral blood

using standard protocols. PolyQ (CAG repeats) and

polyG (GGN repeats) alleles were amplified by poly-

merase chain reaction as previously described [5].

Repeat fragment sizing was performed on an ABI

Prism 3700 DNA Sequencer (Applied Biosystems,

Foster City, CA, USA). The specific length of both

the CAG and GGN repeats was further verified via

Sanger sequencing.

Genomic DNA was extracted from peripheral blood using standard protocols. PolyQ (CAG repeats) and polyG (GGN repeats) alleles were amplified by poly-merase chain reaction as previously described [5]. Repeat fragment sizing was performed on an ABI Prism 3700 DNA Sequencer (Applied Biosystems, Foster City, CA, USA). The specific length of both the CAG and GGN repeats was further verified via Sanger sequencing. Genomic DNA was extracted from

peripheral blood using standard protocols. PolyQ (CAG repeats) and polyG (GGN repeats) alleles were amplified by poly-merase chain reaction as previously described [5]. Repeat fragment sizing was performed on an ABI Prism 3700 DNA Sequencer (Applied Biosystems, Foster City, CA, USA). The specific length of both the CAG and GGN repeats was further verified via Sanger sequencing.

After obtaining informed consent, blood samples were collected. DNA was extracted from EDTA blood using a standard salting-out procedure. The CAG repeat length was calculated as previously described [16].

Haplotypes analysis

Only patients with founders from the Veneto region were considered in the haplotype definition. One proband for family was included in the haplotype analysis and overall 25 patients were analyzed.

To identify AR lineages, we selected a set of 15 highly informative markers, spanning about 2,6 Kb around the AR gene, based on the linkage disequilibrium (LD) block structure of the region and on SNP minor allele frequencies in the European population [17].

We genotyped 7 STR markers (DXS1194, DXS1111, GGN repeats on exon 1 of AR gene [16] and 4 STRs selected with a Tandem Repeats Finder software), the mutated CAG and 7 SNPs (rs34191540, rs6625187, rs17302090, rs7061037, rs2361634, rs1337080, rs12012620) [17].

For STRs markers, PCR amplifications were performed under standard conditions using primers with fluorescent dye (FAM, HEX). DNA fragment length analysis was done with ABI PRISM 3700 DNA Sequencer (Applied Biosystems, Foster City, CA). 4 SNPs were amplified using a standard protocol and genotyped with a restriction enzyme, 2 using a Tetra-primer ARMS PCR and 1 with ARMS PCR. For primers details see Supplementary data (Tab-S1 and Tab-S2).

Statistical analysis

Statistical analysis were performed using JMP13.Pro.

Continuous variables were summarized as mean value, standard deviation and 95% confidence interval (CI), while categorical variables were summarized as frequencies and 95% CI.

Group comparisons of normally distributed variables were performed using t-tests for two groups and one-way ANOVA for more than two groups. For non-normally distributed data, the 2-tailed unpaired Mann-Whitney U test was used. The Chi-squared test was used to compare categorical data. Spearman's correlation coefficient was used to assess correlations. The significance level was set at $p < 0.05$.

Punctual prevalence at January 31st 2018 was calculated as follows:

$$\textbf{Prevalence} = \frac{\text{Number of people affected by the disease}}{\text{Total number of people in the sample}} \times 100.000$$

The CI were calculated using the Poisson distribution.

The mortality rate was calculated as follows:

$$\textbf{Mortality rate:} \frac{\text{Number of deaths within SBMA patients occurring in the studied period}}{\text{Total number of people in the population}} \times 100.000$$

Results

Demographic characteristics of the studied population.

Epidemiological parameters computation was performed on January 31st, 2018.

Sixty-eight SBMA patients from 32 non-related families living in the Veneto region were identified (Table 1).

Table 1. Prevalence of SBMA for the global Veneto region and for single districts in the region calculated on the 31st January 2018.

Mean age of the patients at the end of the study was 57.91 +/- 12.76 years (IC95%: 54.25 - 61.58). Mean age at onset of muscle weakness was 44.76 +/- 12.79 years (IC95%: 40.83 - 48.70), while mean age at genetic diagnosis of the disease was 51.45 +/- 13.99 years (IC 95%: 47.30 - 55.61).

Six out of 68 patients (8%) died during the observation time. Two of them died at old age (> 85 years) for causes not related with SBMA (complications of ischemic heart failure). 3 patients died after respiratory failure secondary to pulmonary embolism (1 patient at age 69) or *ab ingestis* pneumonia (2 patients, respectively at age 71 and 75). One patient died at the age of 67 for a gastric cancer. Mean age of death was 75.5 years +/- 8.19. Mean disease duration until death was 18.33 years +/- 6.62.

The punctual prevalence of the disease on January 31st, 2018 was 1.24/100.000 (IC95% 1.0-1.48) in the general population and 2.58/100.000 (IC95% 1.65-3.35) in the male population. The prevalence of the disease for each single district is described in table 1. As tested by a Chi-squared test, the prevalence of the disease was significantly higher in the districts of Verona, Vicenza and Treviso than in the others ($p < 0.001$).

For an average yearly population of 4.823.069, mortality rate was 0.12/100.000 (IC95% 0.00-0.38) for the general population and 0.25/100.000 (IC95% 0.00-0.59) for the male population.

Haplotypes identification

Haplotypes were studied in 25 patients from independent families. Results are summarized in Table 2.

Three of the analyzed SNPs (rs34191540, rs6625187 and rs12012620) and all STRs were polymorphic and were able to discriminate between different haplotypes, while rs17302090, rs7061037, rs2361634, rs1337080 appeared to be non-informative in our population. The use of STRs in addition to SNPs defined 5 different haplotypes (Table 2). DXS1111 alleles were different even within the same founder haplotype (1 patient in H1 and 2 patients in H2), but they were considered as recent mutational events.

Mean CAG repeats number in the studied families was 46 +/- 3. CAG repeats number was inversely and significantly correlated with age at symptoms' onset ($r = -0.62$, $p < 0.0001$).

Table 2. Haplotypes identified in out cohort of SBMA patients originating from the Veneto region. Alleles defining haplotype are in Italics; CAG mutated alleles are in bold; polymorphisms located on AR gene are highlighted in grey.

Using a one-way ANOVA test, we found that patients carrying haplotype 2 and 5 had significantly longer CAG repeats sequences in the AR gene compared to those carrying haplotype 1 and 3. The result was confirmed both considering only one proband for family ($p=0.0059$) and all the patients ($p < 0.0001$) (Figure 1).

Figure 1. Distribution of mean CAG repeats number according to the different haplotypes described by mean value and standard deviation considering the entire studied population.

Discussion

In this study, we describe the epidemiology of SBMA in the Veneto region in Italy using information obtained from the local "Rare Diseases Registry" [15] and from the reconstruction of the genealogical trees of the affected families. This is the first study dedicated to the epidemiology of SBMA in Italy considering a wide patient population on the basis of reliable data sources. We collected data from 68 SBMA patients and we computed the punctual prevalence of the disease at 1.24/100.000 (IC95% 1.0-1.48) in the general population and 2.58/100.000 (IC95% 1.65-3.35) in the male population.

These results are consistent with those of a previous ascertainment in an adjacent Italian area [11]. However, the estimated prevalence appears to be relevantly lower than that found in the Vasa district in Finland [10]. Neither a different clinical background in terms of disease severity between Italian and Finnish patients, nor relevant differences in the standard of SBMA care, could explain this discrepancy. Rather, patient clustering, caused by the small population considered in the Finland study and the presence of a strong founder effect, may clarify the significant difference observed between the two regions.

Exhaustive information about haplotypes distribution in Italy is not available, even if the presence of several and completely different founders has already been described [12]. In order to evaluate the existence of a common haplotype that could introduce a bias on prevalence estimation, we clarified the ancestral origin of our cohort through a haplotype study, based on slow evolving SNP backgrounds and ancillary STRs spanning 1.5 cM (2.6 Mb) around the AR gene [17]. This study identified 5 different haplotypes in our population. The GGN repeat is located 1.2 kb downstream the pathological CAG, with recombination barely detected between

the two repeats [17]. Thus, the observation in our haplotypes of chromosomes carrying $(GGN)_{20}$, $(GGN)_{23}$, and $(GGN)_{24}$ alleles strongly support a scenario where SBMA mutations arose more than once, according to the ongoing hypothesis of multiple founder effects [14]. H1 and H2 were partially similar and share several markers alleles, including $(GGN)_{20}$, but they differed in the telomeric portion of the haplotype (Table 2). Therefore, we cannot exclude that the two haplotypes may originate from a common ancestor and that a recombination event occurred between rs1337080 and STR4. The identification of different haplotypes suggests that our prevalence estimation is not affected by a strong single founder effect and thus it could represent a reliable estimation of SBMA frequency.

Interestingly, we demonstrated that, within the same haplotype, patients shared a similar CAG repeat number, suggesting that the length of *de novo* mutations is preserved with little variations in the founder's offspring. Literature data on the contraction/expansion of CAG repeats upon parental transmission supported the hypothesis that CAG-SBMA expansion seems relative stable, even if both small expansions and small contractions were observed (involving about 25% of meiosis) [18]. According to these data, our results seem to confirm a substantial stability of AR-CAG expansion within the same lineage.

Mechanisms underlying *de novo* expansions and parental inheritance of the pathogenetic AR-CAG are still not established. Nevertheless, based on data obtained in other polyQ diseases (i.e. SCA3, SCA7 and Huntington disease) [19-21], differences in the *-cis* genetic background could modify meiotic instability of CAG tract. Different haplotypes may influence the parental transmission of expanded repeats length in patients and increase the probability of *de novo* event in normal-range alleles. Further studies considering larger SBMA cohorts and matched control samples, may shed light into the genetic background that influence AR-CAG length through generations, evaluate the existence of large normal or pre-mutated alleles and highlight the mechanisms underlying *de novo* mutations. These information could increase the knowledge about the disease and could be useful to improve parental counselling.

Conflict of interest disclosure

The authors deny any conflict of interest.

Acknowledgements

This study was partially funded by AFM Telethon (to G.S.), ASLA Onlus, Ricerca Finalizzata from the Italian Ministry of

Acknowledgements

This study was partially funded by AFM Telethon (to G.S.), ASLA Onlus, Ricerca Finalizzata from the Italian Ministry of Health (to G.S.), EuroBioBank and Telethon Network of Genetic Biobanks (GTB12001 to EP).

References

1. Querin G, Sorarù G, Pradat PF. Kennedy disease (X-linked recessive bulbospinal neuronopathy): A comprehensive review from pathophysiology to therapy. *Revue Neurologie (Paris)* 2017;173(5):326-337.
1. Manzano R, Sorarú G, Grunseich C, et al. Beyond motor neurons: expanding the clinical spectrum in Kennedy's disease. *Journal of Neurology Neurosurgery and Psychiatry* 2018;89(8):808-812.
2. Sorarù G, D'Ascenzo C, Polo A, et al. Spinal and bulbar muscular atrophy: skeletal muscle pathology in male patients and heterozygous females. *Journal of the Neurological Sciences* 2008;264(1-2):100-105.
3. Querin G, Bertolin C, Da Re E, et al. Non-neural phenotype of spinal and bulbar muscular

atrophy: results from a large cohort of Italian patients. *Journal of Neurology Neurosurgery and Psychiatry* 2016;87(8):810-816.

4. Rosenbohm A, Hirsch S, Volk AE, et al. The metabolic and endocrine characteristics in spinal and bulbar muscular atrophy. *Journal of Neurology* 2018;265(5):1026-1036.
5. Guber RD, Takyar V, Kokkinis A, et al. Nonalcoholic fatty liver disease in spinal and bulbar muscular atrophy. *Neurology* 2017;89(24):2481-2490.
6. Araki A, Katsuno M, Suzuki K, et al. Brugada syndrome in spinal and bulbar muscular atrophy. *Neurology* 2014;82(20):1813-1821.
7. Fratta P, Nirmalananthan N, Masset L, et al. Correlation of clinical and molecular features in spinal bulbar muscular atrophy. *Neurology* 2014;82(23):2077-2084.
8. Finsterer J, Soraru G. Onset Manifestations of Spinal and Bulbar Muscular Atrophy (Kennedy's Disease). *Journal of Molecular Neurosciences* 2016; 58(3):321-329
9. Udd B, Juvonen V, Hakamies L, et al. High prevalence of Kennedy's disease in Western Finland - is the syndrome underdiagnosed? *Acta Neurologica Scandinava* 1998: 98: 128-133.
10. Guidetti D, Sabadini R, Ferlini A, Torrente I. Epidemiological survey of X-linked bulbar and spinal muscular atrophy, or Kennedy disease, in the province of Reggio Emilia, Italy. *European Journal of Epidemiology* 2001;17(6):587-591.
11. Lund A, Udd B, Juvonen V, et al. Founder effect in spinal and bulbar muscular atrophy (SBMA) in Scandinavia. *Eurpean Journal of Human Genetetics* 2000;8(8):631-636.
12. Tanaka F, Doyu M, Ito Y, et al. Founder effect in spinal and bulbar muscular atrophy (SBMA). *Human Molecular Genetics* 1996;5(9):1253-1257.
13. Lund A, Udd B, Juvonen V, et al. Multiple founder effects in spinal and bulbar muscular atrophy (SBMA, Kennedy disease) around the world. *European Journal of Human Genetics* 2001;9(6):431-436.
14. Mazzucato M, Visonà Dalla Pozza L, Manea S, Minichiello C, Facchin P. A population-based registry as a source of health indicators for rare diseases: the ten-year experience of the Veneto Region's rare diseases registry. *Orphanet Journal of Rare Dis*eases 2014; 9:37.
15. Bertolin C, Querin G, Da Re E, et al. No effect of AR polyG polymorphism on spinal and

bulbar muscular atrophy phenotype. *European Journal of Neurology* 2016;23(6):1134-1136.

16. Santos D, Pimenta J, Wong VC, Amorim A, Martins S. Diversity in the androgen receptor CAG repeat has been shaped by a multistep mutational mechanism. America Journal of Medical Genetics Part B Neuropsychiatric Genetics 2014;165B(7):581-6. doi: 10.1002/ajmg.b.32261.
17. Dion V, Wilson JH. Instability and chromatin structure of expanded trinucleotide repeats. *Trends in Genetics* 2009;25(7):288-297.
18. Bingham PM, Scott MO, Wang S, et al. Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice. *Nature Genetica*. 1995;9(2):191-196.
19. Martins S, Coutinho P, Silveira I, et al. Cis-acting factors promoting the CAG intergenerational instability in Machado-Joseph disease. America Journal of Medical Genetics Part B Neuropsychiatric Genetics 2008;147B(4):439-46.
20. Faruq M, Magaña JJ, Suroliya V, et al. A Complete Association of an intronic SNP rs6798742 with Origin of Spinocerebellar Ataxia Type 7-CAG Expansion Loci in the Indian and Mexican Population. Annals of Human Genetics 2017;81(5):197-204. doi: 10.1111/ahg.12200
21. Kay C, Collins JA, Wright GEB, et al.. The molecular epidemiology of Huntington disease is related to intermediate allele frequency and haplotype in the general population. America Journal of Medical Genetics Part B Neuropsychiatric Genetics 2018;177(3):346-357. doi: 10.1002/ajmg.b.32618.

Table 1. Prevalence of SBMA for the global Veneto region and for single districts in the region calculated on the 31st January 2018

District	Number of resident patients	Number of families	Prevalence (IC95%)	Prevalence on male subjects (IC95%)
Verona	25	11	2.71/100.000 (2.36-3.48)	4.65/100.000 (3.68-5.62)
Vicenza	14	7	1.61/100.000 (1.01-2.21)	3.29/100.000 (3.28-3.29)
Venezia	7	1	0.81/100.000 (0.18-1.44)	1.69/100.000 (1.00-2.44)
Treviso	15	10	1.69/100.000 (1.1-2.28)	3.45/100.00 (2.48-4.32)
Belluno	0	/	0.00/100.000	0.00/100.000
Padova	1	1	0.10/100.000 (0.05-0.15)	0.21/100.000 (0.13-0.28)
Rovigo	0	0	0.00/100.000	0.00/100.000
Total	62	30	1.24/100.000 (1.0-1.48)	2.58/100.000 (1.65-3.35)

Table 2. Haplotypes identified in out cohort of SBMA patients originating from the Veneto region. Alleles defining haplotype are in Italics; CAG mutated alleles are in bold; polymorphisms located on AR gene are highlighted in grey.

Haplotype	Proband ID	Geographical origin	DXS1195	rs341915	STR2 ΔAAT	STR3 TTA	rs662518	rs173020	CAG	GGN	rs706103	rs236163	rs133708	STR4 CA	rs120126	STR5 AGG	DXS1111
H1	3778	Verona	*18*	*C*	14	*22*	*A*	G	**46**	*20*	A	A	A	*23*	*G*	*17*	30
H1	3776	Padova	*18*	*C*	14	*22*	*A*	G	**42**	*20*	A	A	A	*23*	*G*	*17*	32
H1	8028	Vicenza	*18*	*C*	14	*22*	*A*	G	**42**	*20*	A	A	A	*23*	*G*	*17*	32
H2	6891	Verona	*18*	*C*	14	*22*	*A*	G	**48**	*20*	A	A	A	*22*	*A*	*19*	30
H2	4526	Verona	*18*	*C*	14	*22*	*A*	G	**47**	*20*	A	A	A	*22*	*A*	*19*	30
H2	7040	Verona	*18*	*C*	14	*22*	*A*	G	**46**	*20*	A	A	A	*22*	*A*	*19*	28
H2	4444	Verona	*18*	*C*	14	*22*	*A*	G	**45**	*20*	A	A	A	*22*	*A*	*19*	28
H3	3396	Treviso	*15*	*T*	14	*22*	*C*	G	**44**	*23*	A	A	A	*24*	*G*	*17*	29
H3	6216	Treviso	*15*	*T*	14	*22*	*C*	G	**42**	*23*	A	A	A	*24*	*G*	*17*	29
H3	7068	Treviso	*15*	*T*	14	*22*	*C*	G	**46**	*23*	A	A	A	*24*	*G*	*17*	29
H3	7422	Treviso	*15*	*T*	14	*22*	*C*	G	**41**	*23*	A	A	A	*24*	*G*	*17*	29
H3	7675	Treviso	*15*	*T*	14	*22*	*C*	G	**46**	*23*	A	A	A	*24*	*G*	*17*	29
H3	7893	Treviso	*15*	*T*	14	*22*	*C*	G	**44**	*23*	A	A	A	*24*	*G*	*17*	29
H3	8025	Treviso	*15*	*T*	14	*22*	*C*	G	**47**	*23*	A	A	A	*24*	*G*	*17*	29
H4	6484	Treviso	*15*	*T*	14	*22*	*A*	G	**44**	*24*	A	A	A	*24*	*G*	*17*	29
H5	3385	Vicenza	*15*	*T*	14	*24*	*C*	G	**45**	*24*	A	A	A	*23*	*G*	*17*	32

H5	4562	Vicenza	*15*	*T*	14	*24*	*C*	G	**5**	**0**	*24*	A	A	A	*23*	*G*	*17*	32
H5	6153	Verona	*15*	*T*	14	*24*	*C*	G	**4**	**8**	*24*	A	A	A	*23*	*G*	*17*	32
H5	6480	Vicenza	*15*	*T*	14	*24*	*C*	G	**5**	**0**	*24*	A	A	A	*23*	*G*	*17*	32
H5	6493	Verona	*15*	*T*	14	*24*	*C*	G	**5**	**2**	*24*	A	A	A	*23*	*G*	*17*	32
H5	6727	Verona	*15*	*T*	14	*24*	*C*	G	**4**	**8**	*24*	A	A	A	*23*	*G*	*17*	32
H5	6751	Verona	*15*	*T*	14	*24*	*C*	G	**4**	**8**	*24*	A	A	A	*23*	*G*	*17*	32
H5	6753	Treviso	*15*	*T*	14	*24*	*C*	G	**4**	**4**	*24*	A	A	A	*23*	*G*	*17*	32
H5	6756	Vicenza	*15*	*T*	14	*24*	*C*	G	**5**	**0**	*24*	A	A	A	*23*	*G*	*17*	32
H5	7004	Venezia	*15*	*T*	14	*24*	*C*	G	**4**	**8**	*24*	A	A	A	*23*	*G*	*17*	32

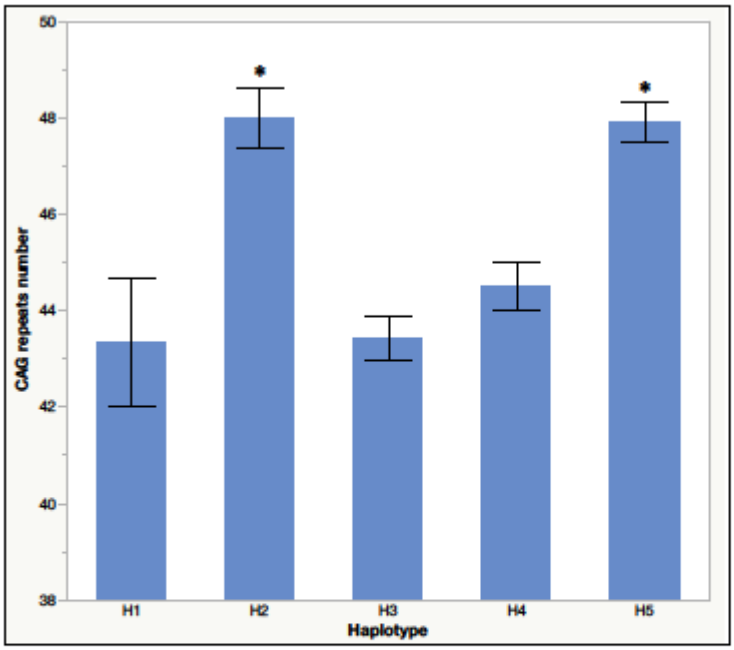

Figure 1. Distribution of mean CAG repeats number according to the different haplotypes described of mean value and standard deviation considering the entire studied population.