Population genetics of spinocerebellar ataxias caused by polyglutamine expansions

A.N. Shuvaev1, O.S. Belozor1, M.V. Smolnikova1, D.A. Yakovleva2, Andr.N. Shuvaev2, O.M. Kazantseva3, E.A. Pozhilenkova1, O.I. Mozhei5, S. Kasparov5, 6

1 Krasnoyarsk State Medical University named after V.F. Voino-Yasenetsky, Research Institute of Molecular Medicine and Pathobiology, Krasnoyarsk, Russia
2 Siberian Federal University, Krasnoyarsk, Russia
3 Krasnoyarsk State Center of Medical Genetics, Krasnoyarsk, Russia
4 Federal Research Center “Krasnoyarsk Science Center” of the Siberian Branch of the Russian Academy of Sciences, Scientific Research Institute of Medical Problems of the North, Krasnoyarsk, Russia
5 Immanuel Kant Baltic Federal University, Kaliningrad, Russia
6 University of Bristol, Bristol, United Kingdom

Hereditary disorders of the neuronal system are some of the most important problems of medicine in the XXI century. The most interesting representatives of this group are highly prevalent polyglutamine spinocerebellar ataxias (SCAs). It has a basement for quick progression of expansion among different groups all over the World. These diseases are SCA1, 2, 3, 6, 7 and 17, which phenotypically belong to one group due to similarities in clinics and genetics. The substrate of these genetic conditions is CAG trinucleotide repeat of Ataxin genes which may expand in the course of reproduction. For this reason a characteristic feature of these diseases is not only an increase in patient numbers, but also a qualitative change in the progression of their neurological symptoms. All these aspects are reflected in the structure of the incidence of polyglutamine SCAs, both at the global level and at the level of individual population groups. However, most scientific reports that describe the population genetics of polyglutamine SCAs are limited to quantitative indicators of a specific condition in a certain area, while the history of the occurrence and principles of the distribution of polyglutamine SCAs are poorly understood. This prevents long-term predictions of the dynamics of the disease and development of strategies for controlling the spread of mutations in the populations. In this paper we make a detailed analysis of the polyglutamine SCAs population genetics, both in the whole world and specifically in the Russian Federation. We note that for a better analysis it would be necessary to cover a wider range of populations in Africa, Asia and South America, which will be possible with the development of new methods for molecular genetics. Development of new methods of detection of polyglutamine SCAs will allow the scientists to better understand how they lead to the brain disease, the means of their spread in the population and to develop better methods for therapy and prevention of these diseases.

Key words: spinocerebellar ataxia; polyglutamine diseases; population genetics; epidemiology.

Population genetics of spinocerebellar ataxias caused by polyglutamine expansions

**Introduction**

In 1991 a new type of mutations in the human genome was discovered, the so-called “dynamic mutations”. These mutations cause an increase in the number of copies (expansion) of the simple repeating sequences (Kremer et al., 1991; Warren 1996). As was revealed later, there are numerous simple (CAG, GCC, GGC, GAA, CTG) (Verkerk et al., 1991; Brook et al., 1992; Matilla et al., 1993; Kawaguchi et al., 1994; Koide et al., 1994; Gedeon et al., 1995; Campuzano et al., 1996; David et al., 1997; Zhuchenko et al., 1997; Babovic-Vuksanovic et al., 1998; Brais et al., 1998; Xiang et al., 1998; Vincent et al., 2000; O’Hearn et al., 2001), and more complex repetitive sequences (CTGG, ATTCT, CCCCGCCCCGCG) (Lalioti et al., 1997; Liquori et al., 2001; Potaman et al., 2003). The most common and heterogeneous group of such diseases is the group caused by expansion of CAG triplet in relevant genes (Table 1). CAG encodes the amino acid glutamine, so these diseases are called “polyglutamine” (Zoghby, Orr, 1999).

Today, we know of at least nine such conditions. The first one, which was described in 1991, linked to CAG expansion and causing the progressive degeneration of motor neurons, was the spinal and bulbar muscular atrophy (SBMA), or Kennedy’s disease (La Spada et al., 1991). Subsequently also this pathological mechanism has been found in eight other diseases, including Huntington’s disease (HD) in the huntingtin gene, dentato-rubral-pallido-luysian atrophy (DRPLA or Haw River syndrome) in the ATN1 gene, and six types of spinocerebellar ataxias (SCA1, 2, 3, 6, 7, and 17) (Gardian et al., 2005).

Among all polyglutamine diseases there, SCAs have the most similar pathophysiology, progression and clinical signs. Their genetics, however, is not identical (see Table 1).

**Features of mutagenesis in polyglutamine SCA progression**

Mechanisms of mutagenic process are dramatically different from those of the static mutations. This explains the dominance of polyglutamine SCAs over SCAs caused by static mutations. In comparison to point mutations, which occur spontaneously and stochastically, dynamic mutations have a substrate, the CAG sequence, which initially is repeated only several times (Dunnen, 2017). CAG sequence expansion usually takes place during mitosis of somatic and germ cells. Trinucleotide repeat expansion occurs by replication-dependent (Kovtun, McMurray, 2001), and repairation-dependent (Kovtun et al., 2007) mechanisms. These disturbances are the cause of the phenomenon called “anticipation”, when the disease occurs progressively earlier and is more severe in subsequent generations. It was discovered that, the longer the allele is, the more unstable it becomes. The average number of CAG repeats expanded during reproduction depends on the SCA type (from +0.5 CAG repeats in SCA3 to +12 CAG repeats in SCA7) (Stevanin et al., 2000). Paternal alleles are more unstable during transmission which is probably due to a larger number of mitotic divisions during sperm cell maturation compared to oocytes during gametogenesis. However, it could also be linked with decreases in repair DNA protein concentration and activity (Pearson et al., 2005).

The total morbidity of polyglutamine SCAs is dramatically variable and varies around 1–9 per 100.000 with more accurate accounts being 4–5 per 100.000. SCA1 is observed, approximately, at a frequency of 1–2 people per 100.000 of “general” population (Manto, 2005). SCA2 has been revealed in 14 % of cases of all SCAs, which makes up about 0.6 per 100.000 population (Cancel et al., 1997; Geschwind et al., 1997a; Riess et al., 1997). SCA3 at some areas is the most common autosomal dominant SCA (Schols et al., 2004; Bauer et al., 2005). The frequency of SCA3 is ~1.5–2 persons per 100.000 population (van de Warrenburg et al., 2002). The world-wide incidence of SCA6 comes near 0.02–0.31 per 100.000 population (Geschwind et al., 1997b; Ikeuchi et al., 1997; Matsumura et al., 1997; Matsuyama et al., 1997; Riess et al., 1997; Stevanin et al., 1997; Schols et al., 1998; Pujana et al., 1999; Jiang et al., 2005). The incidence of SCA7 in several studies was 2 % of all SCAs, but according to the most accurate calculations it is 0.05–0.2 (an average 0.08) per 100.000 (Filla et al., 2000; Storey et al., 2000). Less than 100 families with SCA17 have been described to date (~0.0015 per 100.000) (Maruyama et al., 2002; Alendar et al., 2004; Craig et al., 2005) (Figure, a).

Overall, about 60 % of all clinical SCAs are polyglutamine SCAs, while the other identified and accurately diagnosed forms comprise less than 5 %. It is therefore important to note...
that 35–40% of SCAs have no established genetic base and are only characterized by the phenotype (Bird, 1998; Jayadev, Bird, 2013). At the same time, among polyglutamine SCAs, SCA1 and SCA3 comprise 2/3 of all registered cases (37 and 42% respectively) (see Figure, a).

**Polyglutamine SCAs prevalence rate in Russia**

Research into polyglutamine diseases in Russia has been carried out for over 20 years. The prevalence of polyglutamine SCAs in Russia is similar to those in the European populations, but there are also differences. From 105 Russian families (excluding the Yakut population) with polyglutamine diseases, SCA1, 2, 3, 6, 7 and 17 was diagnosed in 61 families. The prevalence of SCAs are: SCA1, 28 families (46%); SCA2, 21 (34.4%); SCA3, 7 (11.5%); SCA6, 1 family; and SCA7, 1 family (per 1.6%). SCA17 was found in 3 families (Klyushnikov et al., 2016, 2017). Thus, characteristic in the Russian population is a low incidence of SCA3, while in populations of the USA and Japan it is the most frequent pathology (Klyushnikov et al., 2008). Notably, the prevalence of SCA17 is very high in comparison to the distribution in global population.

**Genetic aspects of the origin and expansion of polyglutamine SCAs genesis**

Advancements in population genetics and paleogenetics have allowed scientists to trace when polyglutamine SCAs appeared in the human population. The mutation of a gene is quite an unusual phenomenon, therefore such diseases are characterized by the presence of the founder, the person who for the first time gained a certain mutation and became the origin of the unique genetic profile called “haplotype”. Analyses of haplotypes help reveal the dynamics, common factors that affect the disease and predict the development of this pathology in the future. In polyglutamine SCAs, several haplotypes of founders were traced, with at least 2–3 separate haplotypes in each population (Lund et al., 2001; Bettencourt, Lima, 2011) (see Figure, b and Table 2). Usually, the mutation rate is higher and/or the mutation is older when it can be detected in wider populations or when it can be detected with a higher incidence (Lund et al., 2000; Bettencourt, Lima, 2011). However, due to the “anticipation” phenomenon, these two factors are mutually

**Table 1. Molecular characteristics of polyglutamine SCAs**

<table>
<thead>
<tr>
<th>Disease</th>
<th>Locus</th>
<th>Gene</th>
<th>Protein</th>
<th>Polyglutamine chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCA1</td>
<td>6p23</td>
<td>Atxn1</td>
<td>Ataxin-1</td>
<td>6–39</td>
</tr>
<tr>
<td>SCA2</td>
<td>12q24</td>
<td>Atxn2</td>
<td>Ataxin-2</td>
<td>14–32</td>
</tr>
<tr>
<td>SCA3</td>
<td>14q24-q31</td>
<td>Atxn3</td>
<td>Ataxin-3</td>
<td>12–40</td>
</tr>
<tr>
<td>SCA6</td>
<td>19p13</td>
<td>CACNA1A</td>
<td>α_{m} P/Q Ca^{2+} channel</td>
<td>4–18</td>
</tr>
<tr>
<td>SCA7</td>
<td>3p21–p12</td>
<td>Atxn7</td>
<td>Ataxin-7</td>
<td>7–18</td>
</tr>
<tr>
<td>SCA17</td>
<td>6q27</td>
<td>TBP</td>
<td>TATA-box-binding protein</td>
<td>25–43</td>
</tr>
</tbody>
</table>

Worldwide distribution of polyglutamine SCAs.

a – the pie chart shows the percentage ratio of each polyglutamine SCA relative to the total number calculated per 100.000 population (SCA17 is only 0.037 % of the total, this value is not seen in the graph); b – on the World map is shown the place of occurrence of corresponding disease haplotype (black and white circles).
Population genetics of spinocerebellar ataxias caused by polyglutamine expansions

Table 2. Worldwide distribution of polyglutamine SCAs

<table>
<thead>
<tr>
<th>Mutation rate</th>
<th>SCA</th>
<th>Haplotype</th>
<th>Place of inhabitance</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow (time of origin 80.000 years ago and later)</td>
<td>SCA3</td>
<td>Asiatic</td>
<td>Taiwan, India, Japan, Australia</td>
<td>Gaspar et al., 2001; Verbeek et al., 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Portuguese (global distribution due to the great geographical discoveries)</td>
<td>North America, Germany, France, Portugal, Brazil, India, China, Australia</td>
<td>Bettencourt, Lima, 2011; Martins et al., 2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cambodian (possibly a variation of the Portuguese haplotype)</td>
<td>Cambodia, United States</td>
<td>Jaydev et al., 2006</td>
</tr>
<tr>
<td></td>
<td>SCA6</td>
<td>Paleolithic (before the division of humanity into races)</td>
<td>England, Japan, Brazil, Finland</td>
<td>Craig et al., 2008</td>
</tr>
<tr>
<td>Fast (time of origin 900 years ago and later)</td>
<td>SCA1</td>
<td>Japanese</td>
<td>North Rhine-Westphalia</td>
<td>Dichgans et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Portuguese</td>
<td>North America, Germany, France, Portugal, Brazil, India, China, Australia</td>
<td>Bettencourt, Lima, 2011; Martins et al., 2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cambodian (possibly a variation of the Portuguese haplotype)</td>
<td>Cambodia, United States</td>
<td>Jaydev et al., 2006</td>
</tr>
<tr>
<td></td>
<td>SCA2</td>
<td>Western European</td>
<td>Belgium, Finland, France, Germany, Sweden and UK</td>
<td>Stevanin et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pan-European</td>
<td>France, Germany, Serbia</td>
<td>Didierjean et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>North African</td>
<td>Morocco, Libya</td>
<td>Didierjean et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Caribbean</td>
<td>Jamaica, Cuba</td>
<td>Didierjean et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indian</td>
<td>India (Bihar)</td>
<td>Choudhry et al., 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Japanese</td>
<td>Gunma Prefecture</td>
<td>Mizushima et al., 1998</td>
</tr>
<tr>
<td></td>
<td>SCA7</td>
<td>Haplotype of Continental Europe</td>
<td>Belgium, Finland, France, Germany, Sweden and UK</td>
<td>Stevanin et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Italian</td>
<td>Italy</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asiatic</td>
<td>Korea and Philippines</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Middle Eastern</td>
<td>Israel</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anglo-Saxon</td>
<td>UK, USA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>North African</td>
<td>Algeria, Morocco, Tunisia</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>South American</td>
<td>Brazil</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scandinavian</td>
<td>Sweden</td>
<td>Jonasson et al., 2000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>South African</td>
<td>South African</td>
<td>Greenberg et al., 2006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Japanese</td>
<td>Niigata Prefecture</td>
<td>Koido et al., 1999</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German (high instability of CAG repeats)</td>
<td>Northern Germany</td>
<td>Zühlke et al., 2005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>German (low instability of CAG repeats)</td>
<td>Germany</td>
<td>De Michele et al., 2003; Zühlke et al., 2003</td>
</tr>
</tbody>
</table>

exclusive in polyglutamine SCAs. This conclusion could be suspected by comparing the haplotype numbers and SCA time of origin (see Table 2).

According to the mutation speed, age of clinical presentation and time of origin, polyglutamine SCAs can be divided conditionally into two groups. The first comprises diseases with early presentation and low speed of mutations, SCA3 and SCA6. 2–3 haplotypes are known for each polyglutamine disease.

In the case of SCA6, the size of the area occupied by the diseased population directly correlates with the time of its origin. The worldwide distribution of this pathology charac-
terized by a very low degree of mutation can be linked to the presence of a Paleolithic haplotype which arose before the division of humanity into races (about 80,000 years ago) (Craig et al., 2008). The Asiatic haplotype of SCA3 arose in the prehistoric period (about 7,000 years ago). People with this haplotype live in South and Southeast Asia, and also in northeastern part of Australia (Martins et al., 2012). However, the worldwide distribution of SCA3 is explained differently. The Portuguese haplotype arose ~1400 years ago and was local until the era of great geographical discoveries (Bettencourt, Lima, 2011). The active exploration of the World Ocean by the Portuguese resulted in this haplotype quickly spreading to the countries of Asia and the New World (Martins et al., 2012). While these haplotypes make a significant contribution to the incidence of SCA3 and SCA6, there are also “younger” local haplotypes all over the World (Dichgans et al., 1999; Mori et al., 2001; Jayadev et al., 2006).

The widespread distribution of SCA1, SCA2, SCA7 and SCA17 in the world population is due to a different reason than SCA3, SCA6. These diseases are characterized by a high level of mutations. The mutations that cause these diseases are relatively young: in most cases, they are present only in 1–20 generations and are common within various ethnic groups. Patients with these polyglutamine SCAs exhibit numerous haplotypes, in some cases these are just individual haplotypes due to spontaneous mutagenesis but result from duplications (Zühlke et al., 2003). For this reason, the identification of a haplotype is important not only for assessing the incidence in a single population, but also for predicting the course of the disease and for medical-genetic counseling of patients.

**Isolation as a factor in the incidence of polyglutamine SCAs enhancement**

Isolation is a strong factor contributing to polyglutamine SCAs. Isolation could be seen not only in areas with a low migration history due to the remoteness of the place or ethnic characteristics, but also it could be linked to traditions leading to closely related marriages (Dedov et al., 2004; Maximova et al., 2008). The most vivid example of natural isolation is the population of the small island named Flores of the Azores archipelago, where there is the highest incidence of SCA3 in the world, 1 of 140 people (Lima et al., 1998). Another striking case of natural isolation is the larger Yakut population, where the incidence of SCA1 is about 46 cases per 100 thousand population (Platonov et al., 2016). Until the XX century, the population of the Japanese islands remained in relative isolation from other peoples. Confirmation of this can be found in the official data of SCA1 incidence in Miyagi and Yamagata Prefectures where the migration rate is the lowest in Japan (Wakisaka et al., 1995). However, geographical isolation is always a relative phenomenon. Thus, in China, not only Chinese but also Yakut and Japanese haplotypes have been reported (Zhiou et al., 2001).

However, a high level of genetic isolation of a certain group of people may be due not only to geographical factors. The ethnological (cultural, ethnic etc.) isolation of small populations in India, who obey a cast system, also leads to impressive consequences. Thus, SCA1, which is uncharacteristic of India, is observed in Tamil families living exclusively in two villages of Rajapalayam and Kottamedu of Tamil Nadu state, with an incidence of 1 SCA1 patient per 15 healthy residents (Rengaraj et al., 2005). In general, isolation has no global effect on morbidity (Mori et al., 2001), but it is a decisive factor for a specific group of people who live for a long time in the same territory.

**Conclusions**

Polyglutamine SCAs have appeared a long time ago and are spread all over the World. Insufficent information on polyglutamina SCAs does not allow accurate determination of the incidence and prevalence among some population groups. Extensive regions such as Africa, India, Southeast Asia, remain virtually unexplored in terms of these diseases. However, analysis of available data from population genetics reveals a number of features of polyglutamine SCAs. The common features with other hereditary diseases are the presence of the founder, the dependence of the prevalence on the age of the mutation and the frequency of mutagenesis, as well as the prevalence of SCAs in isolated populations. The specifics of SCAs is determined by their mechanism since they occur not due to spontaneous mutagenesis but result from duplications of pre-existing CAG repeats. This mechanism explains the
prevalence of polyglutamine SCAs over other SCAs in the global population, and also the phenomenon of “anticipation”. The anticipation needs to be taken into account in order to explain the dependence of the time of presentation of the disease on the rate of mutation.

One may speculate that 5–7 thousand years ago mainly diseases with a low mutation rate (SCA3 and SCA6) were present, since types with high rates (SCA1, SCA2, SCA7 and SCA17) lead to a rapid elongation of CAG repeats and the manifestation of the disease at younger ages in subsequent generations, which inevitably leads to the exclusion of the affected individuals from the process of reproduction.

The study of polyglutamine SCAs from the point of view of population genetics makes it possible to better determine their place in the context of genetic diseases of the brain and to understand the biological, ethnic and social features of these diseases. The development of molecular genetics will allow the scientists to cover, in the future, a wide range of populations in Africa, Asia and South America, which might not only change our understanding of the time and pattern of distribution of polyglutamine SCAs, but also reveal new mechanisms of the mutation process and disease progression.

References


Популяционная генетика полиглутаминовых спиноцирекбелярных атаксий

А.Н. Шуваев, О.С. Белозор, М.В. Смольникова …
Е.А. Пожиленкова, О.И. Можей, С. Каспаров
2019
23 • 4

Кавагучи Я., Окамото Т., Таниваки М., Айзава М., Инуэ М., Катасима Я., Ятани М., Сисотон П., Игнатиус Я., Йоханнс Д., Руозе И., Ружел Р., Бруде Т.С. Наследственные атаксии: обзор. Генетика. Медицина. 2013;
Икезури Т., Такано Х., Кобаяси С., Икэути Т., Маруяма М., Ширамо Т., Икэути Т., Масао Т., Сузи С. Наследственная спиноциречебеллярная атаксия типа 7: первое описание в русской популяции. Дегенеративные и Вазальные нарушения нервной системы. 2016;103:39-49. (на русском)
Коида Р., Икезури Т., Оноцо О., Такана Х., Игараси С., Эндо К., Такахаси Х., Кано Р., Икэути Т., Сузи С. Устойчивость к повторному расширению генов CAG в результате репликации в русской популяции. Наследственная спиноциречебеллярная атаксия.

Коида Р., Кобаяси С., Шимохата Т., Икезури Т., Маруяма М., Сэйто М., Ямада М., Такахаси Х., Сузи С. Наследственное расширение генов CAG в результате репликации в результате устойчивого расширения генов CAG в русской популяции. Наследственные спиноциречебеллярные атаксии. 2014;11:9-19. (на русском)


Acknowledgements. The scientific article was written with support of Russian Foundation of Basic Research (RFBR) grant KO_a 17-54-10005 and grant “Competitiveness Enhancement Program” for Immanuel Kant Baltic Federal University.

Conflict of interest. The authors declare no conflict of interest.

Received November 22, 2018. Revised March 19, 2019. Accepted March 22, 2019.