Successful sequencing of Huntington's disease CAG microsatellite orthologs from museum collection specimens

Giulio Formenti[1], Michela Pacifico[1], Raffaele Iennaco[1], Nicoletta Ancona[2], Giorgio Bardelli[3], Michela Podestà[3], Edoardo Razzetti[4], Nicola Saino[5], Chiara Zuccato[1], Elena Cattaneo[1]

[1] Department of Biosciences, Centre for Stem Cell Research, Università degli Studi di Milano and INGM Foundation, [2] Acquario e Civica Stazione Idrobiologica di Milano, [3] Museo di Storia Naturale di Milano, [4] Museo di Storia Naturale, Università degli Studi di Pavia, [5] Department of Biosciences, Università degli Studi di Milano

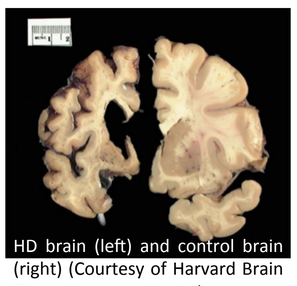
Huntington's Disease is an inherited neurodegenerative disorder affecting human and caused by the expansion of a CAG repeat microsatellite localized in the 5' end of the Huntingtin (Htt) gene, ultimately leading to neuronal loss in brain and motor and neuropsychiatric disturbances. To obtain information on Htt orthologous genes, a potential source of genomic DNA is represented by museum collections. Here we aimed at the characterization of orthologous sequences of the human microsatellite by assessing 114 specimens from Museo di Storia Naturale di Milano, Museo di Storia Naturale di Pavia and Acquario Civico di Milano. Samples belonged to four vertebrate groups: mammals (n=39), amphibians (n=1), fishes (n=22) and reptiles (n=52). Small fragments of tissues preserved in ethanol (n=52), stored frozen (n=79) and tissues that had undergone a tanning process (n=10) were used. These specimens ranged from recently collected to over 50-years-old specimens. Genomic DNA was extracted with two different commercial kits (Macherey-Nagel, Zymo Research) and its quality assessed by ultraviolet-visible spectrophotometry and gel electrophoresis. DNA extraction yielded fragments of highly variable length, probably depending on initial sample preservation conditions. Specific primers designed to encompass the CAG tract and targeting the most conservative regions of the different templates were used to amplify the sequences of interest. Amplicon length ranged from 100 to 250 base pairs. About 60% of the PCR reactions performed were successful and the relative products have been correctly sequenced. This study demonstrates that museum collections are a valuable resource of genomic material to amplify triplet repeat *loci* and restates the importance of specimens conservation for scientific research.

INTRODUCTION

What Huntington's Disease is.

Huntington's Disease (HD) is a neurodegenerative disorder with a fully autosomic dominant inheritance pattern [3, 4]. Prevalence is estimated in one case over 7300 individuals in western population [1, 5, 2]. It was first described in 1872 by George Huntington MD as chorea on the basis of its symptoms characterized by involuntary body movements [9]. In 1993, the responsible gene (IT15 or *HTT*) was identified on chromosome 4 short arm and the encoded protein was named Huntingtin (HTT) [3]. HTT is a highly conserved, ubiquitously expressed protein [8].

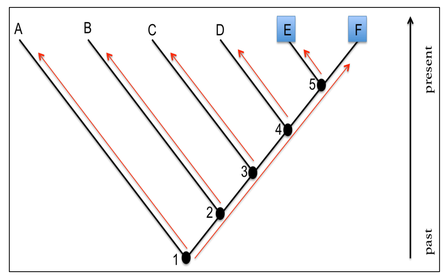
The mutation associated with HD is found in 5' terminal of the HD gene within exon 1 and takes the form of an over-threshold CAG repeats run, which encodes for a poly-Glutamine (polyQ) stretch [3].



HD brain (left) and control brain (right) (Courtesy of Harvard Brain Tissue Resource Center).

A comparative genomics approach.

On the basis of a known phylogeny it is possible to reconstruct what was the ancestral state (*symplesiomorphy*) and the derived state (*synapomorphy*) of a given character [7]. If many species within a group of closely related species share a certain character whileonly one or few of them shows a different state, it is very likely that this last character represents a derived state (i. e. the character that has evolved from the ancestral condition, see Figure). To obtain data for comparative genomics studies a potential source if genomic DNA is represented by museum collections. It is known that DNA molecules under certain conditions can be preserved in tissue samples [6].



SOURCES

Museo di Storia Naturale, Università di Pavia

The Museo di Storia Naturale di Pavia (MSNP), founded in 1771, it is one of the first natural history museums in Europe. Four main collections of high and unique value are present: Spallanzani, Paleontology, Anatomy and Zoology. From Zoology collection we retrieved 52 ethanol preserved specimens of various reptile species (dated 1974 to 2014) and 9 frozen mammals conserved at -20 °C, prior to be processed for taxidermy.

Acquario e Civica Stazione Idrobiologica Milano

The Acquario e Civica Stazione Idrobiologica Milano (ACSIM) was built in 1906 in occasion of the International Exposition of Milano. It's the third most ancient aquarium in Europe. Twenty-two tissues samples from dead fishes of the aquarium conserved at -20°C were provided for our analyses. These included actinopterygians, chondrichthyes and one amphibian and one agnathan.

Museo di Storia Naturale di Milano

The Museo di Storia Naturale di Milano (MSNM) was founded in 1838. Its first collection has grown thanks to donations and with the purchase of finds of priceless scientific value, today is the biggest and most important Italian Museum of natural history. We collected 30 samples (29 mammals and 1 fish). Twenty of them were stored at -20°C, prior to be processed for taxidermy. The others were retrieved from already tanned leather.

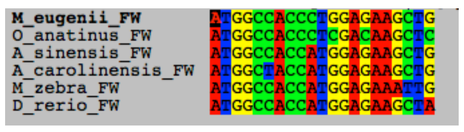
Sources	Order	Family	Species
ACSIM	Cypriniformes	Cyprinidae	*C. auratus*
ACSIM	Perciformes	Acanthuridae	*Z. xanthurum*
ACSIM	Cypriniformes	Cyprinidae	*T. tinca*
ACSIM	Cypriniformes	Cyprinidae	*S. erythrophtalmus*
ACSIM	Perciformes	Carangidae	*T. ovatus*
ACSIM	Perciformes	Apogonidae	*A. imberbis*
ACSIM	Perciformes	Labridae	*T. pavo*
ACSIM	Perciformes	Sparidae	*S. salpa*
ACSIM	Cypriniformes	Cyprinidae	*P. phoxinus*
ACSIM	Perciformes	Sparidae	*D. sargus*
ACSIM	Perciformes	Caproidae	*C. aper*
ACSIM	Siluriformes	Ictaluridae	*A. melas*
ACSIM	Perciformes	Pomacentridae	*C. viridis*
ACSIM	Cypriniformes	Cyprinidae	*T. muticellus*
ACSIM	Cypriniformes	Cyprinidae	*S. cephalus*
ACSIM	Perciformes	Serranidae	*A. anthias*
ACSIM	Salmoniformes	Salmonidae	*T. thymallus*
ACSIM	Salmoniformes	Salmonidae	*S. trutta fario*
ACSIM	Scorpaeniformes	Cottidae	*C. gobio*
ACSIM	Petromuzontiformes	Petromyzontidae	*L. planeri*
ACSIM	Caudata	Ambystomatidae	*A. mexicanum*
ACSIM	Carcharhiniformes	Scyliorhinidae	*S. stellaris*
MSNM	Lamniformes	Lamnidae	*L. nasus*
MSNM	Diprotodontia	Macropodidae	*Macropus sp.*
MSNM	Pholidota	Manidae	*Manis sp.*
MSNM	Artiodactyla	Cervidae	*D. dama*
MSNM	Carnivora	Felidae	*P. tigris*
MSNM	Cingulata	Dasypodidae	*Z. pichiy*
MSNM	Rodentia	Hydrochaeridae	*H. hydrochaeris*
MSNM	Monotremata	Tachyglossidae	*Zaglossus sp.*
MSNM	Rodentia	Bathyergidae	*H. glaber*
MSNM	Rodentia	Ctenodactylidae	*C. gundi*
MSNM	Rodentia	Ctenodactylidae	*P. spekei*
MSNM	Pholidota	Manidae	*Manis sp.*
MSNM	Cetacea	Delphinidae	*S. coeruleoalba*
MSNM	Cetacea	Delphinidae	*S. coeruleoalba*
MSNM	Cetacea	Delphinidae	*S. coeruleoalba*
MSNM	Artiodactyla	Bovidae	*R. rupicapra*
MSNM	Carnivora	Phocidae	*P. vitulina*
MSNM	Rodentia	Sciuridae	*S. vulgaris*
MSNM	Erinaceomorpha	Erinaceidae	*E. europaeus*
MSNM	Carnivora	Mustelidae	*M. nivalis*
MSNM	Lagomorpha	Leporidae	*O. cuniculus*
MSNM	Hyracoidea	Procaviidae	*P. capensis*
MSNM	Carnivora	Mustelidae	*M. putorius*
MSNM	Carnivora	Mustelidae	*M. putorius*
MSNM	Carnivora	Mustelidae	*M. foina*
MSNM	Rodentia	Sciuridae	*M. marmota*
MSNM	Carnivora	Canidae	*V. vulpes*
MSNM	Rodentia	Cricetidae	*M. glareolus*
MSNM	Rodentia	Muridae	*Gerbillus sp.*
MSNM	Cetacea	Balaenopteridae	*B. physalus*
MSNM	Dermoptera	Cynocephalidae	*G. variegatus*
MSNUP	Soricomorpha	Talpidae	*T. europaea*
MSNUP	Soricomorpha	Talpidae	*T. europaea*
MSNUP	Soricomorpha	Talpidae	*T. europaea*
MSNUP	Carnivora	Canidae	*V. vulpes*
MSNUP	Rodentia	Gliridae	*M. avellanarius*
MSNUP	Rodentia	Muridae	*R. norvegicus*
MSNUP	Rodentia	Muridae	*R. norvegicus*
MSNUP	Rodentia	Sciuridae	*S. vulgaris*
MSNUP	Rodentia	Gliridae	*G. glis*
MSNUP	Squamata	Scincidae	*C. chalcides*
MSNUP	Squamata	Scincidae	*C. chalcides*
MSNUP	Squamata	Scincidae	*C. chalcides*
MSNUP	Squamata	Scincidae	*C. chalcides*
MSNUP	Squamata	Scincidae	*C. chalcides*
MSNUP	Squamata	Scincidae	*C. chalcides*
MSNUP	Squamata	Scincidae	*C. chalcides*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*P. tiliguerta*
MSNUP	Squamata	Lacertidae	*P. siculus*
MSNUP	Squamata	Lacertidae	*P. siculus*
MSNUP	Squamata	Lacertidae	*P. siculus*
MSNUP	Squamata	Lacertidae	*P. siculus*
MSNUP	Squamata	Lacertidae	*P. siculus*
MSNUP	Squamata	Lacertidae	*P. siculus*
MSNUP	Squamata	Lacertidae	*P. tiliguerta*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Lacertidae	*A. nigropunctatus*
MSNUP	Squamata	Lacertidae	*A. nigropunctatus*
MSNUP	Squamata	Colubridae	*C. austriaca*
MSNUP	Squamata	Colubridae	*C. austriaca*
MSNUP	Squamata	Colubridae	*C. austriaca*
MSNUP	Squamata	Colubridae	*C. austriaca*
MSNUP	Squamata	Colubridae	*H. viridiflavus*
MSNUP	Squamata	Colubridae	*H. viridiflavus*
MSNUP	Squamata	Lacertidae	*Z. vivipara*
MSNUP	Squamata	Lacertidae	*Z. vivipara*
MSNUP	Squamata	Lacertidae	*P. muralis*
MSNUP	Squamata	Colubridae	*H. viridiflavus*
MSNUP	Squamata	Colubridae	*H. viridiflavus*
MSNUP	Squamata	Natricidae	*N. natrix*
MSNUP	Squamata	Natricidae	*N. natrix*
MSNUP	Squamata	Colubridae	*C. girondica*
MSNUP	Squamata	Colubridae	*C. girondica*
MSNUP	Squamata	Viperidae	*V. aspis*
MSNUP	Squamata	Viperidae	*V. aspis*
MSNUP	Squamata	Lacertidae	*L. bilineata*
MSNUP	Squamata	Lacertidae	*L. bilineata*
MSNUP	Squamata	Anguidae	*A. fragilis*
MSNUP	Squamata	Anguidae	*A. fragilis*
MSNUP	Squamata	Natricidae	*N. tessellata*
MSNUP	Squamata	Natricidae	*N. tessellata*
MSNUP	Squamata	Colubridae	*Z. situla*
MSNUP	Squamata	Colubridae	*Z. situla*
MSNUP	Squamata	Colubridae	*Z. longissimus*

METHODS

Small tissue samples were cut and prepared for DNA extraction. Genomic DNA was extracted with with two different commercial kits (Macherey-Nagel, Zymo Research). gDNA extraction was performed on 114 samples.
We analyzed DNA concentration b NanoDrop 1000 Spectrophotometer (ThermoScientific). Then we performed gel electrophoresis on agarose gel (2%) to asses the DNA degradation.

We tested several PCR reaction protocols in order to develop the best protocol for each vertebrate group. Principal differences were between $MgSO_4$ quantity, annealing temperature and time. Gel electrophoresis was performed using 2-4 uL of PCR product at the end of every reaction.
The amplicon length expected was different for each group of vertebrate. In particular, we expected 100pb for fishes, 150 for reptiles and 200 for mammals. 10 uL of PCR products were used for Sanger sequencing (performed by BMR Genomics, Padova).

Our PCR primers map in the 5' and 3' of HTT exon 1. This is a highly conserved genomic region. Specific primers were designed for each group of to account for the main differences according to the consensus of several known sequences from genome data.



The image shows an example of a few primers designed on known sequences, highlightining the differences among different animals.

RESULTS&DISCUSSION

Genomic DNA from recently frozen samples (fig. 1) have shown the best results whereas ethanol preserved samples (fig. 2) frequently do not preserve the DNA from degradation.

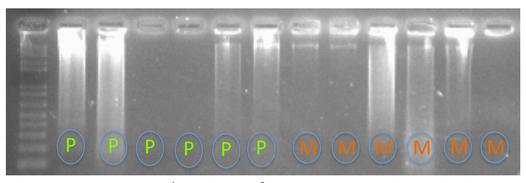


Figure 1. Mammalian DNA, frozen tissues. P = MSNP; M = MSNM.

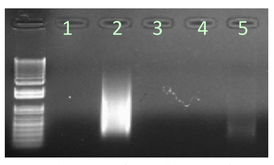


Figure 2. Reptilian DNA, ethanol preserved tissue, MSNP (The numbers correspond to samples in Table 1)

Sample ID	Year	DNA ng/ul	260/280	260/230
1) C. chalcides	2003	5,80	1,12	0,26
2) C. chalcides	2003	326,70	1,90	1,97
3) C. chalcides	1974	8,66	1,22	0,25
4) C. chalcides	1974	1,82	0,89	0,17
5) C. chalcides	1974	7,62	1,62	0,90

Table 1. Year of capture and spectrophotometer results, corrsponding to samples in Figure 2.

In 62% of samples PCR reaction was successful. Success is dependent on DNA quality, the presence of suitable primer and PCR protocols. A PCR was considered "successful" when bands of expected size or of slightly different size was seen. Multiple bands containing the amplicon of expected size were considered successful but protocol optimization is required.

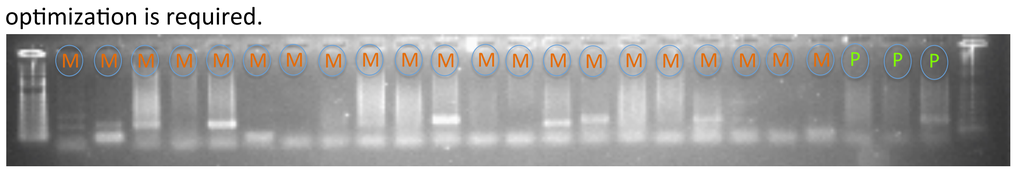


Figure 4. PCR products gel electrophoresis of mammalian samples. P = MSNP; M = MSNM. Expected amplicon length is around 200 pb, in relation to the length of the CAG tract. Ladder 1kb Plus.

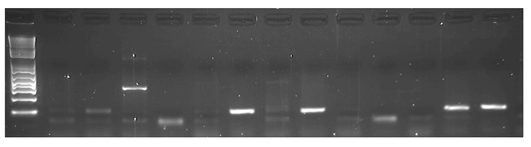


Figure 5. PCR products gel electrophoresis of fish samples (ACSIM). The expected band is at 100 pb. Ladder 1kb Plus.

According to this, we sequenced all the PCR product we obtained. Good quality sequences were produced for the 38,5% of our samples. The discrepancy between the successful PCR and the good quality sequences is due to the presence of nonspecific products that hindered our interest region. In few cases, we observed the failure of sequencing reaction. In most cases, high quality sequences were obtained with primers producing amplicons of at least 150 bp.

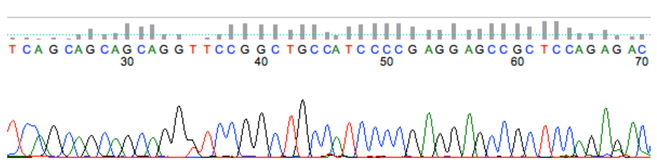


Figure 6. Chromatogram of forward sequencing of A. nigropunctatus

Legenda	Significato
	Processato, ma non produce risultato
	Processato, risultato sia con PCR che sequenziamento, ma con sequenze non di nostro interesse
	Risultato dubbio
	Processato, dati ottenuti positivamente
	Non processato

In our sample, the rapid preservation after death appears to be the most relevant factor in preserving DNA from degradation. Moreover, freezing (-20 C) tends to preserve good quality DNA even when the sample is stored for a long time. On the contrary, conservation in ethanol did not guarantee from DNA degradation, both in recent and old samples. Once the DNA is effectively extracted from the tissue it is generally suitable for PCR amplification and sequencing even when degraded.

Almqvist, EW., et al. «High incidence rate and absent family histories in one quarter of patients newly diagnosed with Huntington disease in British Columbia.» *Clin Genet.*, 2001: 60(3):198-205.
Fisher, ER., et al. «Multisource ascertainment of Huntington disease in Canada: prevalence and population at risk.» *Mov Disord*, 2014: 29(1):105-14.
The Huntington's Disease Collaborative Research Group. «A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes.» *Cell*, 1993: 26;72(6): 971-83.
Lee, JM., et al. «CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion.» *Neurology*, 2012c: 78(10):690-5.
Morrison, PJ., et al. «Uptake of Huntington disease predictive testing in a complete population.» *Clin Genet*, 2011: 80(3):281-6.
Pääbo, S. «Molecular cloning of ancient Egyptian mummy DNA.» *Nature*, 1985: 314(6012):644–645.
Pagel, MD., et al. «The Comparative Method in Evolutionary Biology.» *Oxford Series in Ecology and Evolution*, 1991.
Strong, TV., et al. «Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues.» *Nat Genet*, 1993: 5(3):259-65.
Weiner, W., «Movement disorders: a comprehensive survey.» *Futura Publishing Company, New York*, 1989.