NorthPoleInc.

PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, NORTHPOLE

Heterologous expression and functional characterization of the Santa Hoho2 gene

Santa Claes[1*], Rudolph Reindeer[2], Saint Nicolas[3], Ness E. Tomte[3], Dasher Sridhar[3], Jeremy Elf[1]

[1]*University of NorthPole, Biochemistry Dept., Arctic Lane 1, Santa's Secret Village, NorthPole*

[2]*University of NorthPole, School of Veterinary Medicine, Arctic Lane 25, Santa's Secret Village, NorthPole*

[3]*University of NorthPole, Department of Very Cool Stuff, Freeeezing Cold Avenue 365, Santa's Secret Village, NorthPole*

**Corresponding author: santa@dna20.com*

Received 28 June 2004, received in revised form 25 October 2006, accepted 24 December 2006

Abstract

In this paper we unequivocally identify and characterize the genetic determinant of the famous white beard of Santa Claus to be the ortholog of human KRT6B. The newly discovered gene is named Hoho2 for Human ortholog for hair ougmentation 2. The Santa gene Hoho2 is synthesized and codon optimized for codon expression. Successful heterologous protein expression is shown in three separate systems; *E. coli*, reindeer, and human. We further show that the bearded phenotype is tissue specific in mammalians, but not in prokaryotes. A Hoho2 specific RNAi knockout was constructed and shown to specifically disrupt the facial beard phenotype. Trans-complementation of the gene could be achieved using a synthetic RNAi resistant variant, indicating that the phenotype is truly a direct consequence of the Hoho2 gene and not due to indirect or off-target-effects on the phenotypic display.

keywords: codon optimization, heterologous protein expression, red nose, Hoho2, facial hair, lipid bilayer, transgenic expression

Introduction

The extended and much admired facial hair of Santa Claus was first described by Diedrich Knickerbocker (Knickerbocker, 1809). The density, thickness and majestic splendor of Santa Claus's beard is one of the key identifying features of the true guardian of Christmas values. The phylogenetic origin of the famous beard has been one of the few remaining universal questions looking for an answer in order to validate the existence of Santa Claus. Anthropological studies have further suggested that dense facial hair, extended torso girth phenotype and preference for red clothing are all associated with the early Pleistocene *Homo neanderthalis* culture (Briggs, 2005). Much work and research interest has since been devoted to tracing the evolutionary origin and cultural iconization of the characteristic follicle enhancement.

The environmental challenges associated with living in arctic climate and travel at high altitude requires a degree of body temperature insulation not normally associated with the *Homo sapiens* ecological niche. It has been argued that the Santa Claus beard and insulating lipid formation throughout the torso are retained and extended through the significant and omnipresent evolutionary selection pressure of arctic low temperature and the high wind chill factor aboard the Red-Nose Express. Microarray studies and recent comparative genomics have further identified and characterized unique genetic elements expressed during facial hair development (Pollard and Salama *et al.,* 2006; Uddin and Wildman *et al.*, 2004). In particular, a region located on chromosome 10q11.2 has been shown to retain a high degree of correlation between micro-array data, quantitative trait loci (QTLs) and comparative genomics suggest that this region has evolved exceedingly fast among the north pole population. Several QTLs affecting temporal affinity for late December and social recognition of well behaved children have been mapped to the 10q11.2 region (Alka and Seltzer, 2001). The chromosomal region is also syntenic to regions that have been linked to rosacea in *Rangifer tarandus* (reindeer) (Stroganoff and Lindström, 2004).

The recent DNA sequencing project by Svante Pääbo and coworkers of chromosome 10 derived from *Homo neanderthalis* bones found in the Neander valley, revealed a stretch of ~2,000 bp immediately flanking 10q11.2 where the *Homo neanderthalis* sequence encodes an apparent duplication of KRT6B (human keratin, type II cytoskeletal 6B) (Takahashi, Paladini *et al.*, 1995). The *Homo neanderthalis* gene is distinctly different from the *Homo sapiens* homolog (92% identity). Furthermore, micro-array analysis suggests a 10-fold increase of KRT6B mRNA abundance in the facial epidermis layer *Homo neanderthalis*. The data has been implied to confer a neanderthal origin of the extended facial hair genotype (Green and Krause *et al.*, 2006).

We here synthesize and clone the KRT6B analog from h. *neanderthalis*. The gene is codon optimized and heterologously expressed in human, *E. coli* and *Rangifer tarandus* and shown to induce follicle production in each host. We further validate the phenotype by RNAi knockout of the activity and trans-complementation using an RNAi resistant variant to rescue the phenotype. The gene is hereby named Hoho2 (Human ortholog for hair ougmentation 2).

Materials ands methods

Synthetic genes were designed by and purchased from DNA2.0 (Menlo Park, CA). Growth media and standard molecular biology buffers were purchased from Teknova (Morgan Hill, CA). All protein expression and protein purification work was carried out by Blue Sky Biotech

(Worcester, MA). Restriction enzymes were purchased from New England Biolabs (Ipswish, MA). All materials and chemicals not otherwise noted were purchased from The Elves Workshop Supplies Inc. (Santa's Secret Village, NorthPole) and were of analytical grade or equivalent.

Heterologous expression hosts

E. coli strain Stbl2 (F^- *mcr*A Δ(*mcr*BC-hsdRMS-*mrr*) *rec*A1 *end*A1 *lon gyr*A96 *thi sup*E44 *rel*A1 λ^- Δ(*lac-pro*AB) was used as prokaryotic host for the heterologous expression of Hoho2.

Two eukaryotic systems were also used to validate the phenotype of the protein expression of the Hoho2 gene: A 43 year old human white male volunteer and a reindeer. In both systems, ectodermal application of differentiated embryonic stem cells (*Homo sapiens* or *Rangifer tarandus* derived as applicable) were used as efficient gene delivery vehicles. The synthetic Hoho2 constructs were inserted after hEF1alpha promoter-driven EGFP or human noggin expression vector as previously described (Hwang and Roh *et al.*, 2005; Kim and Do *et al.*, 2005)

Molecular biological procedures

Standard procedures were used for cloning and analysis of DNA, PCR, electroporation, and transformation (Sambrook, Fritsch *et al.*, 1989). All oligonucleotide synthesis and DNA sequencing was performed by DNA2.0 using instruments and consumables from Applied Biosystems (Foster City, CA).

Sequence analysis

DNA sequence tracefiles were analyzed, automatically assembled into full length contiguous sequences and finished using the Staden package (Staden, Judge *et al.*, 2001). Gene DNA sequence design and codon optimization was carried out using Gene Designer (Villalobos and Ness *et al.*, 2006). Homology searches were performed with the Blast algorithm (Altschul and Madden *et al.*, 1997) implemented on a local server.

Protein analysis

E. coli strains harboring the Hoho2 expression plasmids were grown at 37°C in LB with the appropriate antibiotics. Cells from overnight cultures were centrifuged for 10 min at 5,000 × *g* at 4°C. The supernatants were concentrated by precipitation with 80% ammonium sulfate and further purified using a His-tag and Ni-column as previously described (Gupta and Kim *et al.*, 1997). Proteins were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE) followed by Coomassie blue staining.

Genbank accession numbers for the DNA2.0 designed synthetic constructs are: *E. coli* optimized – SANTA001061225, Human optimized – SANTA002061225, Reindeer optimized – SANTA003061225, RNAi resistant – SANTA047061225.

Results and discussion

The DNA sequence locus CAAN01360007 derived from whole genome shotgun sequence from Neanderthal fossil Vi-80, from the Vindija cave in Croatia (Green et al., 2006) encode an open reading frame that has a high BLAST score similarity to the human KRT6B gene. It is however clear that despite the similarities, the *Homo neanderthalis* homolog has evolved from the human homolog through a duplication and subsequent independent evolution. The ratio of synonymous vs. non-synonymous codons (mean Ka:Ks of 0.44) strongly argues for an evolutionary punctuated equilibrium and an adaptive selective pressure constraining the thermo-resilience of the amino acid sequence (Hambrecht and Quist, 2005).

The wt gene was synthesized and cloned behind a set of different promoters. Unfortunately the gene failed to express in any of the several prokaryotic expression systems tested (data not shown).

Codon optimization was subsequently used to ensure that the codon usage of the synthetic gene construct was consistent with the translational constraints of the expression host (Gustafsson, Govindarajan *et al.*, 2004; Stewart and Burgin, 2005). The design of the synthetic construct was made using the available software Gene Designer (freely available from www.DNA20.com). A single synthetic DNA construct having the wt Hoho2 amino acid sequence, but a nucleotide sequence codon optimized for *E. coli* expression was synthesized. The gene was introduced after an N-terminal his-tag and a T7 promoter with a lac operator (all synthesized by DNA2.0) and subsequently cloned into a vector with a pBR322 based replicon. Results from protein expression can be seen in fig 1.

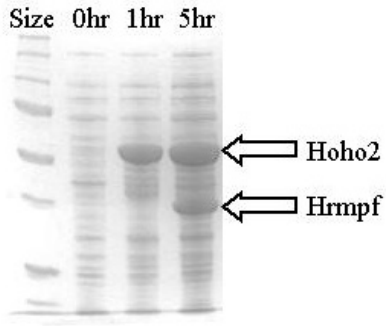

Fig.1. SDS-PAGE of *E. coli* crude extract at three time points after IPTG induction of protein expression: 0hr, Before IPTG induction; 1hr, 1 hour after IPTG induction; 5hr, 5 hours after IPTG induction. The expected size of Hoho2 is marked with arrow. Also marked is Hrmpf (Hoho2 related molecular protein factor).

The heterologous *E. coli* expression system produces large amounts of a protein of the same size as the calculated size of Hoho2. We further identified the presence of a Hoho2-related molecular protein factor (Hrmpf) upon extended incubation during inducing conditions. We have some experimental evidence suggesting that the Hrmpf is an autocatalytically cleaved subdomain of Hoho2. The proteolytic activity is induced over time, but can also be induced by EtOH, lipids, carbohydrates and children yelling at a high pitch (data not shown).

Colonies of *E. coli* strains expressing the codon optimized Hoho2 were shown to have a somewhat rugged morphology when growing on rich media plates. The phenotypic effects of the Hoho2 gene were further analyzed by electron microscopy. As can be seen in figure 2, the entire bacterium is completely covered by extended growth of facial hair. We believe that prokaryotic systems lack the necessary cellular differentiation to separate facial tissue from non-facial tissue. Expression of Hoho2 leads to stimulate follicle activity and correlated production of protruding KRT6B-related polymers.

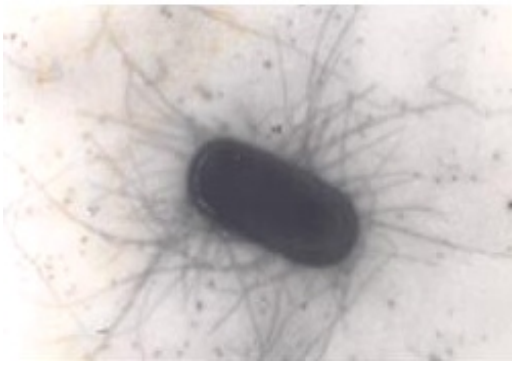

Fig.2. *E. coli* Stbl2 with plasmid pXmas encoding the Hoho2 gene behind a lac repressor. Single cells were isolated 1 hr after IPTG induction and subjected to electron microscopy as previously described (Coca and Cola, 2001).

In order to validate the functional trans-complementation follicle inducing activity of the synthetic Hoho2 gene in a eukaryotic system, we re-synthesized the gene this time optimized for expression in *Rangifer tarandus*. The synthetic gene was

inserted behind the CMV promoter in a hygromycin resistant vector and transformation a *Rangifer tarandus* stem-cell line to create a Hoho2 transgenic *Rangifer tarandus* individual as previously described (Pranser, Dasher *et al*., 2003). As can be seen in fig 3, chin hair follicle induction is rampant on the animal and the reindeer display significant presence of beard formation.

In reindeer embryos, the transgenic Hoho2 is expressed in the periderm (the outermost layer of embryonic epidermis). A subset of periderm cells, localized to temporary epithelial fusions, is shown to contain the keratin 6 orthologous protein, and we find that these cells also harbor 25-30 fold increased follicle activity as measured *in vivo* with reindeer preantral follicle cultures.

Fig.3. *R. tarandus* Xmas06 with plasmid pKkkkold_Hoho2 encoding the Hoho2 gene behind a CMV promoter. Large white keratin formations can be observed immediately below the mouth of this Hoho2 transgenic individual. The beard formation appears upon reaching reindeer puberty (~18 months) and coincides with sleeping in late on weekends.

The *in vivo* follicle assay using preantral follicles from prepubertal female reindeer further identifies direct effects of environmental chemicals such as ginger, marzipan and glühwein on the somatic compartment, the follicle and theca cells, that lead to rapid lipid accumulation in the hepatopancreatic ampulla, vermiform appendix and esophagus (Manuscript in preparation, Dasher et al.,).

As a final corroborative evidence for the functional phenotype of the KRT6B ortholog Hoho2 encoding the classic and characteristic majestic white beard of Santa Claus, a transgenic Hoho2 *Homo sapiens* having beard^{+} phenotype was created. We further showed that the beard^{+} phenotype could be suppressed by an RNAi knockout directed at the Hoho2 gene. In addition, the RNAi induced beard^{-} phenotype could be rescued by trans-complementation using a synthetic RNAi resistant Hoho2 homolog as an antisupressor. The RNAi resistant Hoho2 gene variant was designed following the established RNAi control against off-target-effects (OTE) first validated in plants (Kumar, Gustafsson *et al.*, 2006). The synthetic RNAi resistant gene was designed using the Gene Designer codon optimization 'as-distant-as-possible' algorithm (Villalobos et al., 2006).

Complementation by expression of the RNAi resistant Hoho2 variant indicates that silencing of wt Hoho2 is primarily, if not totally, responsible for the beard^{-} phenotype. It also argues that our syn gene complementation approach is a useful way to address OTE of siRNA. Contrary to the microarray approach to validation, this approach directly assesses the phenotype resulting from RNAi and is therefore independent of the mechanism(s) for silencing the target gene or off-target gene(s). Moreover, it enables the researcher to quantify the contribution of OTE to the silenced phenotype when this phenotype is only partially rescued by the syn gene.

In a separate set of experiments to be published in an upcoming review by Elf et. al., a study of population genetics with

samples derived from the local inhabitants of Burträsk, Sweden, showed increased KRT6B expression in foot sole tissue. It is tempting to speculate this being the origin of the legendary hairy feet of Burträsk.

Fig.4. Phenotypic analysis of the presence of facial keratinated polymers were analyzed in a quadruplicate set of clonal *Homo sapiens* generated from an anonymous and healthy male volunteer as described in materials and methods. **Frame A**, Empty vector control sample. No detectable facial follicle activity as determined by bioscan. **Frame B**, Vector with heterologously expressed Hoho2 from a tissue specific CMV promoter. Large white keratin formations can be observed on chin and lower parts of cheeks. **Frame C.** Same as frame B but with the knockout RNAi (Hoho2 + knockout RNAi) abolishing the beard$^+$ phenotype. **Frame D,** same as frame C but with a trans-complementing RNAi resistant Hoho2 (Hoho2 + knockout RNAi + RNAi resistant version of Hoho2, restoring the beard$^+$ phenotype.

In this paper we unequivocally identify and characterize the genetic determinant of the famous white beard of Santa Claus to be the KRT6B ortholog Hoho2. The genetic distribution of heterologous expression indicate a wide evolutionary presence of auxiliary keratinaceous factors ranging from prokaryotes to mammalians, suggesting that the concept of Santa Claus' features is an historical remnant from before the phylogenetic divergence between prokaryotes and eukaryotes.

We are further studying the Hoho2 transgenic Homo sapiens quadruplicates generated in the process of this paper to assay other Santa Claus related features such as seasonal temperamental variations, affinity to chimneys and affection to red clothing. We expect to report a follow-up study on these matter in due time.

Acknowledgements

The preantral follicles were surgically removed from the pre-pubertal female reindeer by Drs Olesiuk, tenBosch and Wang. We sincerely appreciate their carving skills. Assistance with raising the reindeer litter came from Drs. Villalobos, Spofforth and Pimentel. The initial experimental outline was proposed by Dr. B. Cormack at John Hopkins School of Medicine. Financial support for the work came from the DNA2.0 Research Council grant #001.

References

ALKA, J., and SELTZER, M. (2001). Mapping of various things for various reasons. *Hum. Genet.* **32**(33)**,** 33-44.

ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W., and LIPMAN, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17)**,** 3389-402.

BRIGGS, J. B. (2005). America's deadliest home videos: Rednecks and other neanderthals. *Joe-Bob's Drive-In* **666**(2)**,** 4-56.

COCA, C., and COLA, P. (2001). Gardening is the art that uses flowers and plants as paint, and the soil and sky as canvas. *J. of Electron Microscopy* **23**(3)**,** 66-78.

GREEN, R. E., KRAUSE, J., PTAK, S. E., BRIGGS, A., RONAN, M. T., SIMONS, J. F., DU, L., EGHOLM, M., ROTHBERG, J. M., PAUNOVIC, M., and PÄÄBO, S. (2006). Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**(7117)**,** 330-336.

GUPTA, G., KIM, J., YANG, L., STURLEY, S. L., and DANZIGER, R. S. (1997). Expression and purification of soluble, active heterodimeric guanylyl cyclase from baculovirus. *Protein Expr Purif.* **10**(3)**,** 325-30.

GUSTAFSSON, C., GOVINDARAJAN, S., and MINSHULL, J. (2004). Codon bias and heterologous protein expression. *Trends Biotechnol.* **22**(7)**,** 346-53.

HAMBRECHT, W., and QUIST, G. (2005). Physics investigates the essential nature of the world, biology describes a local bump. *J. of Random Stuff* **65**(34)**,** 1223-65.

HWANG, W. S., ROH, S. I., LEE, B. C., and KANG, S. K. (2005). Patient-specific embryonic stem cells derived from human SCNT blastocysts. *Science*.

KIM, J. H., DO, H. J., CHOI, S. J., CHO, H. J., and PARK, K. H. (2005). Efficient gene delivery in differentiated human embryonic stem cells. *Exp Mol Med.* **37**(1)**,** 36-44.

KNICKERBOCKER, D. (1809). History of New York. *New England Journal of History* **1**(1)**,** 1-25.

KUMAR, D., GUSTAFSSON, C., and KLESSIG, D. F. (2006). Validation of RNAi Silencing Specificity Using Synthetic Genes: Salicylic Acid-binding Protein 2 Is Required For Plant Innate Immunity. *Plant J.* **45**(5)**,** 863-8.

POLLARD, K. S., SALAMA, S. R., LAMBERT, N., LAMBOT, M., IGEL, H., ARES, M., VANDERHAEGHEN, P., and HAUSSLER, D. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**(7108)**,** 167-72.

PRANSER, N., DASHER, K., and DONDER, T. (2003). Use of the cytomegalovirus promoter for transient and stable transgene expression in reindeer embryonic stem cells. *Proc Natl Acad Sci NorthPole* **45**(22)**,** 67-77.

SAMBROOK, J., FRITSCH, E. F., and MANIATIS, T. (1989). "Molecular cloning. A laboratory manual." 2:nd edition ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

STADEN, R., JUDGE, D. P., and BONFIELD, J. K. (2001). Sequence assembly and finishing methods. *Methods Biochem Anal.* **43,** 303-22.

STEWART, L., and BURGIN, A. B. (2005). Whole gene synthesis: A Gene-O-Matic future. *Frontiers in Drug Design & Discovery* **1,** 297-341.

STROGANOFF, S., and LINDSTRÖM, K. (2004). Laser treatment of rosacea in reindeer: a genetic study. *Arch Dermatol.* **140**(11)**,** 1345-9.

TAKAHASHI, K., PALADINI, R. D., and COULOMBE, P. A. (1995). Cloning and characterization of multiple human genes and cDNAs encoding highly related type II keratin 6 isoforms. *J Biol Chem.* **270**(31)**,** 8581-92.

UDDIN, M., WILDMAN, D. E., JOHNSON, R. M., HOF, P. R., KAPATOS, G., GROSSMAN, L. I., and GOODMAN, M. (2004). Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. *Proc Natl Acad Sci U S A.* **101**(9)**,** 2957-62.

VILLALOBOS, A., NESS, J. E., GUSTAFSSON, C., MINSHULL, J., and GOVINDARAJAN, S. (2006). Gene Designer: A synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7,** 285.

WISH YOU A HAPPY HOLIDAY AND A GREAT 2007