ISSN 1925-430X (Print)
ISSN 1925-4318 (Online)

Journal of Molecular Biology Research

CANADIAN CENTER OF SCIENCE AND EDUCATION®

Vol. 5, No. 1 December 2015

Editorial Board

Contents

New Advances Reconstructing the Y Chromosome Haplotype of Napoléon the First Based on Three of his Living Descendants 1
Gérard Lucotte & Peter Hrechdakian

Identifying Significant Biological Markers in Klotho Gene Variants Across Wide Ranging Taxonomy 11
Tommy Rodriguez

Gene Expression Patterns in Functionally Different Cochlear Compartments of the Newborn Rat 20
Johann Gross, Heidi Olze & Birgit Mazurek

αS1-Casein Lineage Assessed by RFLP in the Endangered Goat Breed "Retinta Extremeña" 32
José L. Fernández-García & María P. Vivas Cedillo

Studies on Genetic Diversity of Selected Population of Hybrid Scallop *Chlamys farreri*（♀）× *Patinopecten yessoensis*（♂）by Microsatellites Markers 39
Biao Wu, Aiguo Yang, Ningning Cheng, Xiujun Sun, Zhihong Liu & Liqing Zhou

Real Time PCR: the Use of Reference Genes and Essential Rules Required to Obtain Normalisation Data Reliable to Quantitative Gene Expression 45
Antônio J. Rocha, José E. Monteiro-Júnior, José E.C. Freire, Antônio J.S. Sousa & Cristiane S.R. Fonteles

Study on the Heterosis of the First Generation of Hybrid between Chinese and Korean Populations of S*capharca broughtonii* using Methylation-Sensitive Amplification Polymorphism (MSAP) 56
Hailin Sun, Yanxin Zheng, Chunnuan Zhao, Tao Yu & Jianguo Lin

Reviewer Acknowledgements for Journal of Molecular Biology Research, Vol. 5, No. 1 65
Bella Dong

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

New Advances Reconstructing the Y Chromosome Haplotype of Napoléon the First Based on Three of his Living Descendants

Gérard Lucotte[1] & Peter Hrechdakian[2]

[1] Institute of Molecular Anthropology, 44 Monge Street, Paris 75 005, France

[2] Unifert Group S.A., 54 Louise Avenue, Immeuble Stéphanie, Bruxelles 1050, Belgium

Correspondence: Gérard Lucotte, Institute of Molecular Anthropology, 44 Monge Street, Paris 75 005, France. E-mail: lucotte@hotmail.com

Received: September 5, 2014 Accepted: September 19, 2014 Online Published: December 19, 2014

doi:10.5539/jmbr.v5n1p1 URL: http://dx.doi.org/10.5539/jmbr.v5n1p1

Abstract

This paper describes the findings of the complete reconstruction of the lineage Y chromosome haplotype of the French Emperor Napoléon I. In a previous study (Lucotte et al., 2013) we reconstructed, for more than one hundred Y-STRs (Y–short tandem repeats), the Y-chromosome haplotype of Napoléon I based on data comparing STR allelic values obtained from the DNA of two of his living descendants: Charles Napoléon (C.N.) and Alexandre Colonna Walewski (A.C.W.); in the present study we compare STR allelic values of C.N. and A.C.W. to those of Mike Clovis (M.C.), a living fifth generation descendant of Lucien (one of Napoléon's brothers). When compared between M.C., C.N. and A.C.W., STR allelic values are identical for a total of 93 STRs; that permits us to propose those values, for which the three living descendants are identical, as expected allelic values of Napoléon I's Y-chromosome haplotype. For seven STRs, allele values are variable between M.C., C.N. and A.C.W.; we propose for three of them (DYS442, DYS454 and DYS712) expected allelic values, based on data concerning the allele distributions of these STRs in the population.

Keywords: Napoléon the First, lineage reconstruction, Y-chromosome haplotype

1. Introduction

The French Emperor Napoléon the First (1769-1821), was the son of Charles Bonaparte (1746-1785) and Letizia Ramolino (1750-1836). He had four brothers: Joseph (1768-1844), Lucien (1775-1840), Louis (1778-1846) and Jérôme (1784-1860).

In a first study (Lucotte et al., 2011) we determined the Y-chromosome non-recombinant part (NRY) haplogroup of Napoléon, based on genomic DNA extracted from two islands of follicular sheats associated with his beard hairs conserved in the Vivant-Denon reliquary (Lucotte, 2010). This haplogroup, established by the study of 10 NRY-SNPs (single nucleotide polymorphisms), is E1b1b1c1*; an "oriental" haplogroup of origin, as shown by the frequency map of M34 in contemporary European populations (Lucotte and Diéterlen, 2014), the antepenultimate SNP of the E1b1b1c1* differentiation.

In this same first study (Lucotte et al., 2011) we studied the buccal smear DNA of Charles Napoléon (C.N.), the living fourth generation of male descent from Jérôme, for the first 37 NRY-STRs (short tandem repeats) of the Family Tree DNA (FT DNA) kit; that permits us to establish a first Y-STR profile of C.N. This profile is highly indicative of the E1b1b1 haplogroup, because of STR allelic values at the discriminant (from Athey, 2006) Y-markers DYS19 (allele *13*) and at DYS464.a, .b,.c and .d (alleles *13*, *14*, *15* and *16* respectively); moreover allele values (of *13*) at DYS19 and at YCaII.a and .b (*19* and *22*) are the same for Napoléon (N) and for C.N.

In a second study (Lucotte et al., 2013) we established a more complete (because based on the FTDNA-111STRs kit) Y-STR profile of C.N., and the 111-Y-STRs profile of Alexandre Colonna Walewski (A.C.W.), the fifth generation descendant of Alexandre Walewski (1810-1868) who was the son born of the union between Napoléon I and Countess Maria Walewska (1786-1817). Comparisons at the time between the two STRs profiles were realized for a total number of 130 STRs, six of them (DYS454; DYS481; DYS635 = Y-GATA-C4; DYS712; DYS724 = CDY.a and DYF397.2) having different allelic values between C.N. and A.C.W. At that time we only had three direct determinations available on real allele values of Napoléon (for DYS19 = *13*, and for the palindromic YCAII.a= *19* and .b = *22*).

We then proposed (Lucotte et al., 2013) a first reconstruction of the Y-chromosome haplotype of Napoléon, based on the expected STR allelic values obtained from the 124 identical STRs between C.N. and A.C.W.

We have obtained now (Lucotte & Bouin Wilkinson, 2014) sixteen supplementary allelic direct determinations (on a lock of hair dandruff dating from 1811) for Napoléon I STRs, in order; DYS 19 = *13*; palindromic DYS385.a and b. = *16*; DYS389.i = *14*, .ii = *31*; DYS390 = *24*; DYS391 = *10*; DYS392 = *11*; DYS393 = *14*; DYS438 = *10*; DYS439 = Y-GATA-A4 = *12*; DYS448 = *20*; DYS456 = *15*; DYS458 = *16*; Y-GATA-C4 = *23* and Y-GATA-H4 = *11*. These results confirm our previous ones for allele *13* at DYS19; moreover all these other direct determinations (except for Y-GATA-C4) are in accordance with our previous direct predictions (Lucotte et al., 2013) concerning the expected values for the corresponding STRs.

Mike Clovis (M.C.) is the fifth generation descendant (Figure 1) of Lucien; to visualize the generations of the two male descent from the Walewski and the Jérôme lines, see the first figure of the Lucotte et al., 2013 article. In order to realize a triangular comparison between three living males related to Napoléon I (a direct descendant: A.C.W.; an indirect descendant from his brother Jérôme: C.N.; and M.C., an indirect descendant from his brother Lucien), we study now in the present article the Y-STR profile of M.C. by means of the FT-DNA – 111 STRs kit; and we compare this STR profile to those of C.N. and A.C.W.

2. Methods

Mike Clovis (M.C.) is the *propositus* (Figure 1) for this study. Buccal swab samples for this DNA donor were collected with informed consent. DNA extraction and STRs typing ("upgrade" for 111 genetic markers) were done according to FTDNA recommendations.

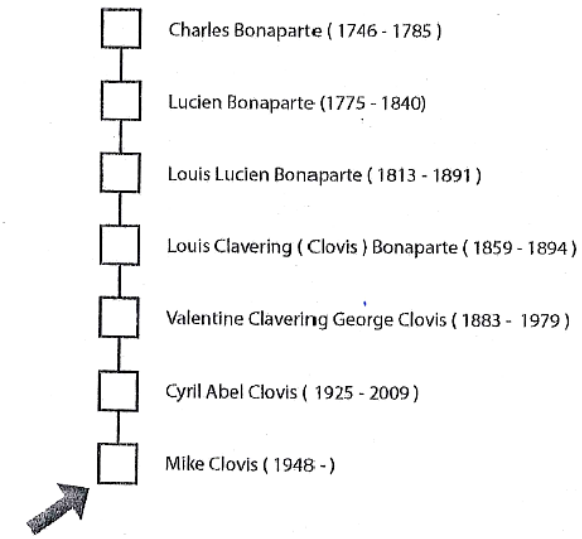

Figure 1. Chain of transmission (seven successive generations of paternal ancestry) from the ancestor Charles Bonaparte (Napoléon's father) to the *propositus* (arrow) Mike Clovis

3. Results

3.1 Comparisons of STR Allelic Values Between M.C. and C.N. – A.C.W.

Table 1 compares, for a total number of 106 STRs, allelic values obtained for M.C. to those of Charles Napoléon (C.N.) and Alexandre Colonna Walewsky (A.C.W.).

Seven STRs show different alleles between these three individuals, in order: DYS442 with an allele value = *11* for M.C. compared to *12* for both C.N. and A.C.W.; DYS447 with an allele value = *22* for M.C. compared to *21* for both C.N. and A.C.W.; for DYS454 the allele value = *11* for M.C. is identical to that of A.C.W., C.N. having an allele value = *7*; for DYS481 the allele value = *27* for M.C. is identical to that of C.N., A.C.W. having an allele value = *28*; for Y-GATA-C4 the allele value = *22* for M.C. is identical to that of C.N., A.C.W. having an allele value = *21*; for DYS712 the allele value = *23* for M.C. is identical to that of C.N., A.C.W. having an allele value = *25*; and for the palindromic CDY.a the allele value of M.C. = *34* is identical to that of A.C.W., C.N. having an allele value = *35*.

Table 1. Allelic values at 106 Y-STRs (numbers refer to the Y-markers of the FTDNA 111 "upgrade" for Mike Clovis (M.C.), Charles Napoléon (C.N.) and Alexandre Colonna Walewski (A.C.W.) NRY-DNAs. Asterisks indicate the seven differential markers (in italics) between M.C., and C.N. and A.C.W.

Numbers	Y-STRs	Allelic values					
		Napoléon I	M.C.	C.N.	A.C.W.	Napoléon I	expected
3	DYS19 = DYS394	13	13	13	13	13	(direct determination)
5	DYS385.a (palindromic)	16	16	16	16	16	(direct determination)
6	.b	16	16	16	16	16	(direct determination)
8	DYS388		12	12	12	12	
10	DYS389.i	14	14	14	14	14	(direct determination)
12	.ii	31	31	31	31	31	(direct determination)
2	DYS390 = DYS708	24	24	24	24	24	(direct determination)
4	DYS391	10	10	10	10	10	(direct determination)
11	DYS392 (located in the untranslated region of the transcription unit TTTY10)	11	11	11	11	11	
1	DYS393= DYS395	14	14	14	14	14	(direct determination)
49	DYS413.a (palindromic)		22	22	22	22	
50	.b		22	22	22	22	
48	DYS425 (one copy of DYF371)		0	0	0	0	
7	DYS426 = DYS483		11	11	11	11	
105	DYS434		9	9	9	9	
53	DYS436		12	12	12	12	
19	DYS437= DYS457		14	14	14	14	
37	DYS438 (located in the untranslated region of the USP9 Y gene)	10	10	10	10	10	(direct determination)
9	DYS439 = Y-GATA-A4		12	12	12	12	(direct determination)
89	DYS441		14	14	14	14	
36	DYS442*		*11*	12	12	12	
57	DYS444= DYS542		11	11	11	11	
86	DYS445		11	11	11	11	
60	DYS446		12	12	12	12	
18	DYS447*		*22*	21	21	21?	
20	DYS448 (located in the P3 loop)	20	20	20	20	20	(direct determination)
21	DYS449		28	28	28	28	
56	DYS450		7	7	7	7	
85	DYS452		30	30	30	30	
17	DYS454*= DYS639		11	7	11	11	
16	DYS455 (located in the intron 2 of the TBL1 Y gene)		11	11	11	11	
30	DYS456	15	15	15	15	15	(direct determination)
13	DYS458	16	16	16	16	16	(direct determination)

14	DYS459.a (palindromic)	9	9	9	9
15	.b	9	9	9	9
26	DYS460= Y-GATA-A7.1	10	10	10	10
106	DYS461= Y-GATA-A7.2	11	11	11	11
84	DYS462	12	12	12	12
88	DYS463	18	18	18	18
22	DYS464.a (palindromic)	14	14	14	14
23	.b	15	15	15	15
24	.c	16	16	16	16
25	.d	17	17	17	17
45	DYS472	8	8	8	8
58	DYS481*	*27*	*27*	*28*	*28* or 27?
69	DYS485	15	15	15	15
63	DYS487= DYS698	14	14	14	14
54	DYS490	12	12	12	12
66	DYS492= DYS604	10	10	10	10
80	DYS494	9	9	9	9
71	DYS495	15	15	15	15
103	DYS497	14	14	14	14
96	DYS504= DYS660	16	16	16	16
75	DYS505	13	13	13	13
104	DYS510	17	17	17	17
47	DYS511	10	10	10	10
97	DYS513= DYS605	12	12	12	12
59	DYS520= DYS654	18	18	18	18
79	DYS522	12	12	12	12
38	DYS531= DYS600	10	10	10	10
94	DYS532	11	11	11	11
81	DYS533	11	11	11	11
55	DYS534	15	15	15	15
43	DYS537	12	12	12	12
72	DYS540	11	11	11	11
77	DYS549	12	12	12	12
76	DYS556	12	12	12	12
51	DYS557	21	21	21	21
98	DYS561	15	15	15	15
67	DYS565	11	11	11	11
62	DYS568	12	12	12	12
33	DYS570 (located in the untranslated region of the TBL1 Y gene)	19	19	19	19
64	DYS572	11	11	11	11

83	DYS575		8	8	8	8	
32	DYS576		18	18	18	18	
39	DYS578		8	8	8	8	
101	DYS587		22	22	22	22	
78	DYS589		11	11	11	11	
42	DYS589		7	7	7	7	
92	DYS593		16	16	16	16	
52	DYS594		11	11	11	11	
31	DYS607		12	12	12	12	
61	DYS617		13	13	13	13	
70	DYS632		8	8	8	8	
100	DYS635*= Y-GATA-C4	23	22	22	21	23	(direct determination)
82	DYS636		11	11	11	11	
65	DYS640= DYS606		13	13	13	13	
44	DYS641		11	11	11	11	
102	DYS643		12	12	12	12	
93	DYS650		18	18	18	18	
68	DYS710		31	31	31	31	
91	DYS712*		23	23	*25*	23	
73	DYS714		24	24	24	24	
95	DYS715		23	23	23	23	
74	DYS716		28	28	28	28	
34	DYS724 = CDY.a* (palindromic) = gene		34	35	34	35 ?	
35	.b		36	36	36	36	
99	DYSS726		15	15	15	15	
87	Y-GATA-A10		12	12	12	12	
27	Y-GATA-H4	11	11	11	11	11	(direct determination)
90	Y-GGATT-1B07		13	13	13	13	
28	YCAII.a (palindromic)	19	19	19	19	19	(direct determination)
29	.b		22	22	22	22	(direct determination)
40	DYF395S1.a (palindromic)	22	15	15	15	15	
41	.b		15	15	15	15	
46	DYF406S1		10	10	10	10	

Because you know exactly how many generations ago the ancestor lived (Figure 1), it is interesting to see how statistics / probabilities compare with reality. In this particular case: Mike > Cyril > Valentine-Louis Clavering > Louis > Lucien > Charles Bonaparte (= Carlo Buonaparte), the probabilities based on the calculations of the time of the most recent common ancestor (the TMRCA) calculations (from Walsh, 2001) are incorrect: comparing the STRs showing only 6 mismatches, it only estimates the probability that Mike Clovis (328303) and Alexandre Colonna Walewski (218983) shared a common ancestor within the last 1 generation = 0.07%, 2 generations = 1.13%, 3 generations = 4.93%, 4 generations = 12.48%, 5 generations = 23.38%, 6 generations = 36.19%, 7 generations = 49.27%,…; so, about 36 to 50% probability six generations back.

3.2. STRs with Identical Values

A total number of 82 STRs have identical allelic values between M.C., C.N. and A.C.W.: in order, the 71 non-palindromic STRs DYS388 = *12*; DYS425 = *0*; DYS426 = *11*; DYS434 = *9*; DYS436 = *12*; DYS437 = *14*; DYS441 = *14*; DYS444 = *11*; DYS445 = *11*; DYS446 = *12*; DYS449 = *28*; DYS450 = *7*; DYS452 = *30*; DYS455 = *11*; DYS460 = *10*; DYS461 = *11*; DYS462 = *12*; DYS463 = *18*; DYS472=*8*; DYS485 = *15*; DYS487 =*14*; DYS490 = *12*; DYS492 = *10*;DYS494 = *9*;DYS495 = *15*;DYS497 = *14*; DYS504 = *16*; DYS505 = *13*; DYS510 = *17*;DYS511 = *10*; DYS513 = *12*; DYS520 = *18*; DYS522 = *12*; DYS531 = *10*; DYS532 = *11*; DYS533 = *11*; DYS534 = *15*; DYS537 = *12*; DYS540 = *11*; DYS549 = *12*; DYS556 = *12*; DYS557 = *21*; DYS561 = *15*; DYS565 = *11*; DYS568 = *12*; DYS570 = *19*; DYS572 = *11*; DYS575 = *8*;DYS576 = *18*; DYS578 = *8*; DYS587 = *22*; DYS589 = *11*; DYS590 = *7*; DYS593 = *16*; DYS594 = *11*; DYS607 = *12*; DYS617 = *13*; DYS632 = *8*; DYS636 = *11*; DYS640 = *13*; DYS641 = *11*; DYS643 = *12*; DYS650 = *18*; DYS710 = *31*; DYS714 = *24*; DYS715 = *23*; DYS716 = *28*; DYS726 = *15*; Y-GATA-A10 = *12*; Y-GGAAT-1B07 = *13* and DYF40651=*10*.

Likewise for the 11 palindromic STRs: DYS413.a = *22*,.b = *22*; DYS459.a = *9*,.b=*9*; DYS464.a=*14*,.b=*15*,. c=*16*,. d=*17*, CDY.b=*36* and DYF395S1.a=*15*,.b =*15*,which have identical values between M.C., C.N. and A.C.W.

Because of this identity, we can reasonably infer that the 93 allelic values of the above genetic markers correspond to those expected for Napoléon I (because they have remained unchanged for 5/6 generations of remote ancestry).

3.3 Differential STRs

Table 2 lists and characterizes the seven STRs that differentiate between M.C., C.N., and A.C.W. Only one of them (CDY.a) is palindromic. The mutation rates, when known (Burgarella & Navascués, 2011), of these differential alleles are in the 10^{-3} range (except for DYS447). These rates are impossible to evaluate for the palindromic STR CDY. a, and unknown for the moment for DYS712.

Table 2. Allele values for the seven differencial Y-STRs between Mike Clovis (M.C.), Charles Napoléon (C.N.) and Alexandre Colonna Walewski (A.C.W.). Expected allele values for **N** (Napoléon I) are established for the three Y-STRs DYS454 = *11*, DYS712 = *23* and DYS442 = *12*

Numbers	**Y-STRs**	**Palindromic**	**Mutation rates**	**N** (direct determination)	**Allele values**			**Racial background**	**N** (deduced / expected values)
					M.C.	C.N.	A.C.W.		
1	**Y-GATA-C4**		2.832x 10-3	23	22	22	21		
2	**DYS454**		$2.182x10^{-3}$		11	7	11		**11**
3	**DYS712**		?		23	23	25		**23**
4	**DYS442**		$1.926x10^{-3}$		11	12	12		**12**
5	**DYS447**		$7.414x10^{-4}$		22	21	21	+?	**21?**
6	**DYS481**		$6.937x10^{-3}$		27	27	28	+	**27-28?**
7	**CDY.a**	+			34	35	34		**35?**

We already know (Lucotte & Bouin-Wilkinson, 2014) the real allele value = *23* of Y-GATA-C4 for Napoléon I. Figure 2 shows the bimodal distribution of Y-GATA-C4 alleles in the population; Napoléon I (N) value corresponds to that of the second modal class. Allele values (=*22*) of Charles Napoléon (C.N.) and Mike Clovis (M.C.) can be explained admitting one-step (*minus* 1) mutations, and that (=*21*) of Alexandre Colonna Walewsky (A.C.W.) admitting a two-step (*minus* 2) mutation.

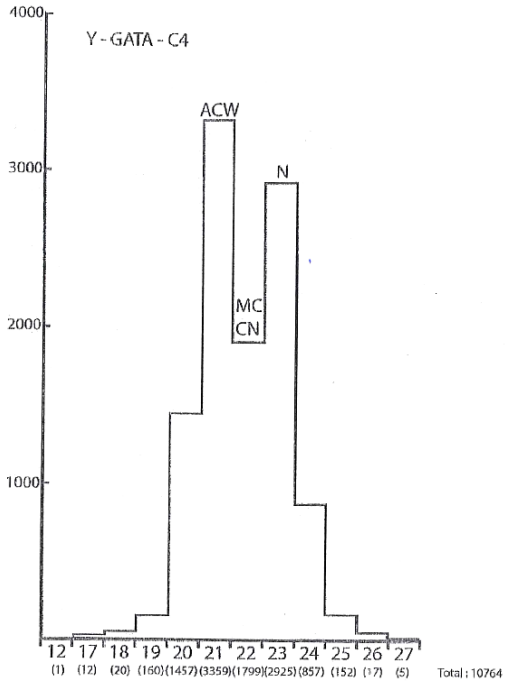

Figure 2. Y-GATA-C4 allelic distribution (www.genebase.com/in/dnaMarkerDetail.php?t=y&d=DYS635), based on values from 10 764 subjects

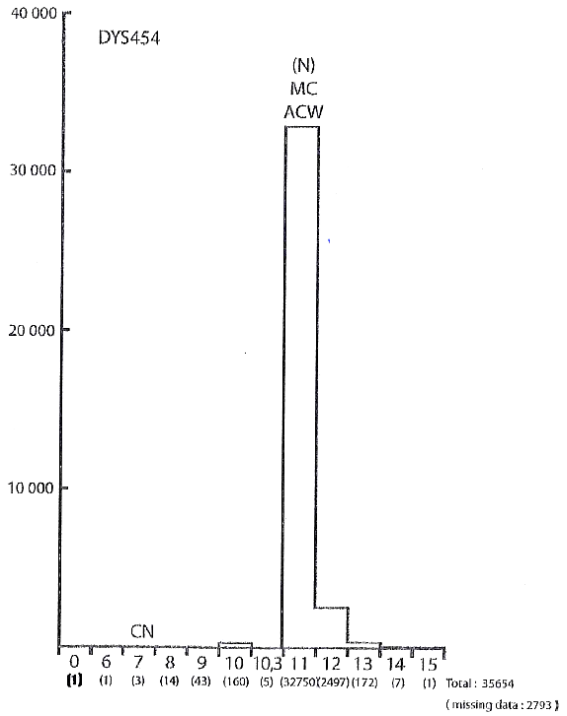

Figure 3. DYS454 allelic distribution (CEPH database), based on values from 35 654 subjects

The distribution of DYS454 alleles in the population is shown in Figure 3. According to Redd *et al.* (2002) DYS454 is one of the most stable (with a pre-eminent modal class = *11*) of the marker set. Because both A.C.W. and M.C. alleles belong to this modal class, the most parsimonious interpretation is that allele *11* at this marker is the ancestral form - that of Napoléon I (N) - and that the allele value = *7* for C.N. represents a derived one, which happened during one of the five generations separating C.N. from the common ancestor Carlo Buonaparte. It is probable that this variant *7* (characteristic of the Jérôme line) is due to a multistep deletion, a rare event which often results in a most stable allele (Lucotte et al., 2013).

Although based on a relatively low number of subjects studied, figure 4 shows a representative allele distribution for DYS712. The interest of this recently described marker is that it certainly represents one of the most variable STR of the panel (its modal class corresponds to allele *22*). Alleles of C.N. and M.C. = *23* and the A.C.W. allele = *25*. The expected N value is probably *23*, because this class corresponds to the second one in importance; and the A.C.W. value = *25* must be due to a two-step (*plus* 2) mutation.

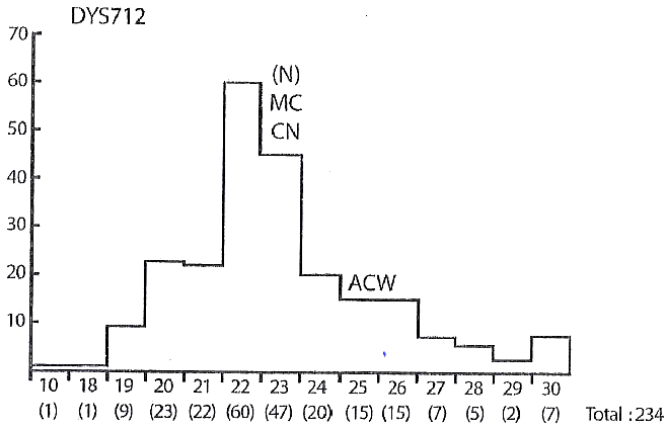

Figure 4. DYS712 allelic distribution (K. Norved, personal communication) of 234 American Caucasians

Figure 5 shows the modal distributions of alleles for DYS442 and DYS447, based on a sample of 1000 European subjects (English: 56, Germans: 59, French Parisians: 191, French Basques: 97, Corsicans: 328, North Italians: 46, Central-Italians: 112, Sardinians: 111; from Diéterlen and Lucotte, 2005). For DYS442 the allele value = *12* of C.N. and A.C.W. corresponds to the modal class; so it is probable that the expected N value is *12*. The value = *11* for M.C. must correspond to a one-step (*minus* 1) mutation.

The pattern of variations is more complicated to interpret for DYS447, because none of the allele values is located at the modal (=*25*) nor at the adjacents (*26* and *24*) classes: allele values for C.N. and A.C.W. = *21*, and *22* for M.C. We presume that the DYS447 distribution presents a small peak in frequencies (possibly due to racial background) at the left tail of the modal class. In this hypothesis, but it is highly speculative, the expected N value could be = *21* (because of the two *21* allelic values of C.N. and A.C.W.).

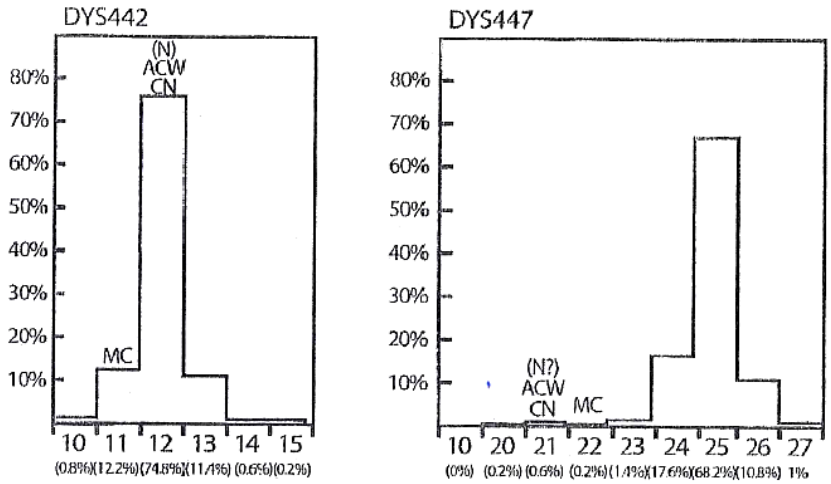

Figure 5. DYS442 and DYS447 allelic distributions, based on our sample of 1 000 unrelated European Caucasians

The existence of this racial background is evident for DYS481 (English: 102, Indians: 83, Africa: 94; from D'Amato et al., 2010), where the three European, Asiatic and African distributions are superposed on the graph (Figure 6): the *27* (for C.N. and M.C.) and *28* (for A.C.W.) classes are relatively well represented at the right tail of the Asiatic distribution, but none (for *27*) or few (for *28*) for the European distribution; but in any case we cannot decide if the expected N value is *27* or *28*.

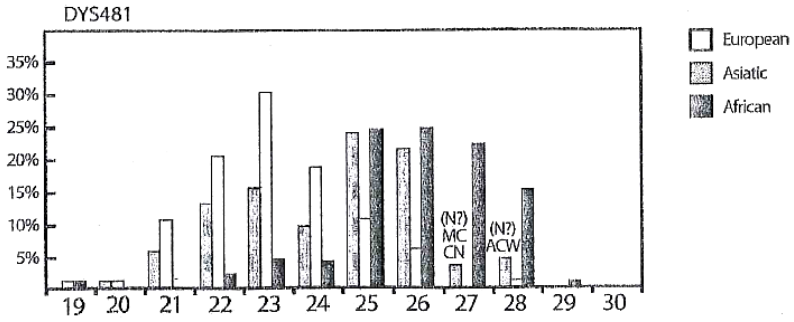

Figure 6. DYS481 allelic distribution (from d' Amato et al., based on samples of 102 European subjects (English), 83 Indians and 94 Xhosas

This sort of racial context intervening for some STR alleles determining the Y-haplotype is interesting to consider, because of the oriental origin (Lucotte & Diéterlen, 2014) of the E1b1b1 haplogroup of Napoléon I, more precisely known now (Lucotte et al., 2013) as E1b1b1b2a1 L792^{+} haplogroup.

Figure 7 shows the modal distributions – based on our sample of 1000 European subjects – of allelic classes for the palindromic markers CDY.a and CDY.b. It is because of the identity of allele values = *36* between C.N., A.C.W.

and M.C. that we proposed that the expected N value for CDY.b is *36*. (though it corresponds to the fourth frequency class in importance, located at the left tail of the modal class).

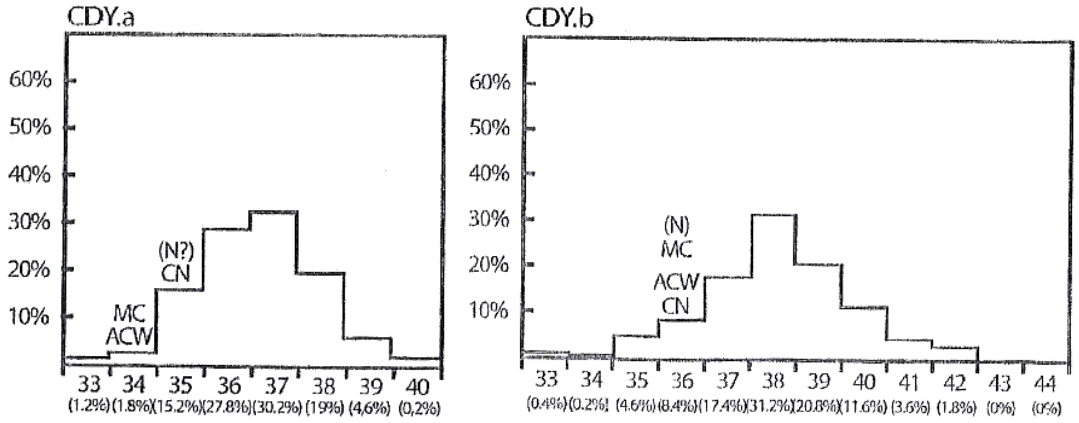

Figure 7. CDY.a and .b allelic distributions, based on the sample of 1 000 unrelated European Caucasians

Predictions about the variation of palindromic Y-STRs, even in the more simple situation of a two-copy marker like CDY, is a very hazardous matter (Lucotte et al., 2013) because we do not know exactly the precise mechanisms involved. For CDY.a the C.N. allelic value = *35*, and the A.C.W. and M.C. values = *34*. We retain here the possibility, but it is also highly speculative, that the expected N value could be = *37* (because, as for CDY.b, it corresponds to the third frequence class, located at the left tail of the modal one). Certainly the *35* alleles for A.C.W. and M.C. cannot be explained by such a simple mechanism as that of one-step (*minus* 1) mutation.

4. Discussion

In the goal to establish the Y-chromosome haplotype of Napoléon the First we determined initially, in his genomic DNA extracted from two islands of follicular sheats associated with his beard hairs conserved in the Vivant-Denon reliquary (Lucotte et al., 2011), allelic values for the three Y-STRs DYS19 and for two palindromic STRs YcaII.a and .b. Subsequently (Lucotte and Bouin-Wilkinson, 2014), based on genomic DNA of his hair dandruff dating from 1811, we determined allelic values for 16 STRs: DYS19 (for which we confirmed the first allelic value previously obtained), the palindromic STRs DYS385.a and .b, DYS389.i and ii, DYS390, DYS391, DYS392, DYS393, DYS438, the variable STR-Y-GATA-A4, DYS448, DYS456, DYS458, Y-GATA-C4 and Y-GATA-H4. The corresponding allele values for these 18 STRs correspond to the real allelic values of the Napoléon I Y-haplotype.

Because of the identity of allelic values of STRs between Charles Napoléon (the living fourth generation of male descent from Jérôme (Napoléon I's youngest brother) and Alexandre Colonna Walewski (a direct living sixth generation descendant from Napoléon I), we proposed (Lucotte et al., 2013) expected allelic values of Napoléon I for a total number of 109 STRs (33 of them being palindromics). For some of the six variables (between Charles and Alexandre) STRs: DYS454, DYS481, Y-GATA-C4, DYS712, CDY.a (palindromic) and DYF397.2 (palindromic), we proposed as expected allelic values for Napoléon I the most probable allelic forms according to STR distributions; the allele value of DYS454 = *7* for Charles Napoléon appeared then as a highly discordant one.

Mike Clovis is a living, previously unknown, fifth descendant of Lucien (another brother of Napoléon I). The objective of the present study is to compare, for a total number of 106 STRs, allelic values between him, Charles Napoléon and Alexandre Colonna Walewski. Identity of allelic values between the three was confirmed for 82 non-palindromic STRs and for 11 palindromic STRs; that confirms, in a triangular form, that these 93 STR allelic values are definitely those previously proposed as expected allele values of the Napoléon I Y-haplotype.

These comparisons between Mike Clovis, Charles Napoléon and Alexandre Colonna Walewsky permit us to clarify some of the questions asked by the variable values between them: for DYS454, the allele value = *11* for Mike Clovis is the expected allelic value of Napoléon I, as previously proposed. For DYS712, the allele value = *23* for Mike Clovis corresponds also to the expected allelic value of Napoléon I already proposed; however in this case it is not the modal class of distribution of DYS71 values that is concerned, but the nearest one of the right edge of this distribution.

Compared to Charles Napoléon and Alexandre Colonna Walewski, Mike Clovis had different allele values for DYS442 = *11* and DYS447 = *22*. For DYS442, as proposed previously, allele value = *12* is probably the expected allelic value of Napoléon I because it corresponds to the modal class of the distribution; and allele value = *11* for Mike Clovis results from a single mutational event (one-step, *minus* 1).

It is impossible to predict some expected allelic value of Napoléon I for DYS447, because the three obtained allele values (that of Mike Clovis = *22* could be the result of a one-step *plus* 1 mutational event) are all located at the left tail of the distribution. It is impossible also to predict some expected allelic value of Napoléon I for DYS481, even when interpreted in the context of the oriental origin of the E1b1b1c1 haplogroup (Lucotte and Diéterlen, 2014), because all the three obtained allele values are now located at the right tail of the distribution.

We ignore, for the moment, what is the Y-chromosome haplotype of Carlo Buonaparte; but it seems highly probable, because of the similarities between the Y-STR values presently obtained, that he was the biological father of Lucien, Napoléon and Jérôme (all these three having the same Y-haplogroup). As a by-product of such studies, we established that the allele value = *7* for DYS545 is highly characteristic of the Jérôme line; possibly, as shown here, the allele value = *11* for DYS442 could be characteristic of the Lucien line. It remains a possibility that allele values of *25* for DYS712 and of *28* for DYS481 could be characteristics of the direct Napoléon I line, at least for the Walewski descent.

References

Athey, W. T. (2006). Haplogroup prediction from Y-STR values using a Bayesian allele frequency approach. *J. Genet. Geneal., 2*, 34-39.

Burgarella, C., & Navascués, M. (2011). Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur. J. Hum. Genet., 19*, 70-75. http://dx.doi.org/10.1038/ejhg.2010.154

D'Amato, M. E., Ehrenreich, L., Cloete, K.,Benjeddou, M., & Davison, S. (2010). Characterization of the highly discriminatory loci DYS449, DYS481, DYS518, DYS612, DYS626, DYS644 and DYS710. *Forens. Sci. Int. Genet., 4*, 104-110. http://dx.doi.org/10.1016/j.fsigen.2009.06.011

Diéterlen, F., & Lucotte, G. (2005). Haplotype XV of the Y-chromosome is the main haplotype in West-Europe. *Biomed. Pharmacother, 59*, 269-272. http://dx.doi.org/10.1016/j.biopha.2004.08.023

Lucotte G., & Bouin-Wilkinson, A. (2014). An autosomal STR profile of Napoléon the First. *Op. J. Genet., 4*, 292-299. http://dx.doi.org/10.4236/ojgen.2014.44027

Lucotte, G. (2010). A rare variant of mtDNA HSV1 sequence in the hairs of Napoléon's family. *Invest. Genet., 1*, 1-4. http://dx.doi.org/10.1186/2041-2223-1-7

Lucotte, G., & Diéterlen, F. (2014). Frequencies of M34, the ultimate genetic marker of the terminal differenciation of Napoléon the First's Y-chromosome haplogroup E1b1b1c1 in Europe, Northern Africa and the Near East. *Int. J. Anthropol., 29*(1-2), 27-41

Lucotte, G., Macé, J., & Hrechdakian, P. (2013). Reconstruction of the lineage Y chromosome haplotype of Napoléon the First. *Int. J. Sciences., 9*, 127-139.

Lucotte, G., Thomasset, T. & Hrechdakian, P. (2011). Haplogroup of the Y chromosome of Napoléon the First. *J. Mol. Biol. Res., 1*, 12-19. http://dx.doi.org/10.5539/jmbr.v1n1p12

Redd, A. J., Agellon, Al B., Kearney, V. A., Contreras, V. A., Karafet, T., Park, H., … Hammer, M. F. (2002). Forensic value of 14 novel STRs on the human Y chromosome. *Forens. Sci. Int., 130*, 97-111. http://dx.doi.org/10.1016/S0379-0738(02)00347-X

Walsh, B. (2001). Estimating the time of the most recent ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics, 158*, 897-912.

Copyrights

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

Identifying Significant Biological Markers in Klotho Gene Variants Across Wide Ranging Taxonomy

Tommy Rodriguez[1]

[1] Pangaea Biosciences, Department of Research & Development, Miami, FL, USA

Correspondence: Tommy Rodriguez, Pangaea Biosciences, 1020 North Lakemont Ave., Winter Park, FL 32792, USA. Tel: 1-305-389-4156. E-mail: trodriguez@pangaeabio.com

Received: November 4, 2014 Accepted: December 7, 2014 Online Published: January 9, 2015

doi:10.5539/jmbr.v5n1p11 URL: http://dx.doi.org/10.5539/jmbr.v5n1p11

Abstract

Biological aging is marked by progressively degenerative physiological change that causes damage to tissues and organs. Errors in biopolymers accumulate over time; mitochondrial dysfunction, telomere attrition, and wider genomic instability lead to an altered state of intercellular communication. In this investigation, my focus will be aimed at examining and identifying specifically critical biomarkers in genetic variants of KLOTHO (a transmembrane protein involved in the genetic regulation of age-related disease) among organisms with varied life spans that range across wide taxonomical rankings. Here, I investigate the correlation between lower and higher frequency α-amino acid compositions in Klotho protein factors within a grouped methodology; as to also include several demonstrative techniques in comparative sequence analysis for inferring relatedness in evolutionary context.

Keywords: Klotho, KL gene, protein, protein evolution, comparative sequence analysis, multiple sequence alignment, Kalign, phylogenetc analysis, amino acid composition analysis

1. Introduction

Promising new research in life extension has led to remarkable advancements in our basic understanding of the molecular mechanisms associated with aging. Within the last two decades, much of this research has focused on identifying potential genomic candidates for longevity, while seeking to explain how these individual genes can affect the biological process of aging. One of the proteins in particular (KLOTHO) has been the subject of several recently published papers. A handful of studies now suggest that genetic variants of KLOTHO [encoded by the KL gene] are associated with human aging and tumor suppression, and trials on model organisms involving KLOTHO variants has shown improvement in cognition and deceleration in age-related development (Dubal et al., 2014). In one such case, Dubal et al. (2014) demonstrated that systemic overexpression of Klotho variants in transgenic mice enhanced cognition and increased longevity by an average ratio of ±25 percent; whereas Klotho-deficient mice manifested a syndrome resembling accelerated human aging and displayed extensive and accelerated arteriosclerosis (Dubal et al., 2014).

For reasons that are not yet fully understood, Klotho-associated mechanisms, “change cellular calcium homeostasis, by both increasing the expression and activity of TRPV5 and decreasing that of TRPC6 (Kurosu et al., 2005).” Moreover, altered mineral-ion homeostasis could be a cause of premature aging‐like phenotypes (Kurosu et al., 2005). In order to gain a more comprehensive understanding of the underlying functions in molecular components of KLOTHO, we should begin by examining KL-derivative patterns across a wide evolutionary spectrum, without limiting ourselves to one individual taxa or another. Because the lifespan of organisms vary widely among species, a comparative approach could help us identify a set of unique signatures in the molecular variation patterns of KLOTHO. In turn, this may help provide meaningful reference toward a fully systematic investigation. Throughout the course of this paper, I seek to address those areas thoroughly.

It should then be noted: This paper does not presume on a solution to longevity; nor does it seek to draw parallels between the many collective biological features that allow organisms to persist in semi-optimal states. Instead, my focus will be aimed at identifying and examining specifically critical biomarkers, or sequence motifs, in genetic variants of KLOTHO among organisms with varied life spans that range across wide taxonomical rankings. Here, I investigate the correlation between lower and higher frequency α-amino acid compositions in Klotho protein factors within a grouped methodology; as to also include several demonstrative techniques in comparative sequence analysis for inferring relatedness in evolutionary context.

2. Methods

2.1 Sequence Selection and Group Categorization

Biological aging is marked by progressively degenerative physiological change that causes [irreversible] damage to tissues and organs. Errors in biopolymers accumulate over time; mitochondrial dysfunction, telomere attrition, and wider genomic instability lead to an altered state of intercellular communication. Within a broader spectrum, degrees of longevity vary widely among organisms. Only a tiny percentage of species exhibit characteristics that is counter-intuitive to universal mechanisms of biological aging. *Turritopsis dohrnii* and *Polycelis feline* are two such examples of eukaryotic organisms that appear to exhibit, "limitless telomere regenerative capacity fueled by a population of highly proliferative adult stem cells (Tan et al., 2012)." Several other species of marine Crustaceans share similar traits that prolong biological aging.

Individual-scale longevity is largely determined by inheritance; genes can explain up to 35 percent (American Federation for Aging Research, 2012). As noted earlier, recent studies indicate that Klotho gene variants may be potential candidates for the genetic regulation of age-related disease. For this purpose, I selected the KL gene as a prime candidate for analysis and evaluation. Furthermore, using homologous variants of Klotho protein sequences as a primary source for comparative studies, I grouped and categorized six distinct genomic datasets of KLOTHO [in FASTA format] by the following criteria:

i. Group *type a* { *Heterocephalus glaber*; *Chelonia mydas*; *Danio rerio*; } – consisting of organisms that are recognized as having favorable biological attributes in one or more of the following areas: life cycle duration, tumor-suppression, cancer-resistance, immunity, tissue regeneration.

ii. Group *type b* { *Homo sapiens*; *Mus musculus*; *Mesocricetus auratus*; } – consisting of organisms that reside within a moderate range of natural life expectancy due to age-related decline, disease, and predispositions.

The six protein sequences used for curating the experimental dataset were obtained via NCBI Protein database. The accession references for each individual sequence are detailed in the sections below.

2.1.1 Regarding group *type a*

Two of three *type a* organisms exhibit extraordinarily prolonged life cycles. *Heterocephalus glaber* are highly resistant to cancer, and can maintain a youthful vascular function and cellular oxidant-antioxidant phenotype relatively longer and are better protected against aging-induced oxidative stress than shorter-living rodents (Buffenstein, 2009; Csiszar et al., 2007; Pérez et al., 2009). *Danio rerio* have shown the ability to regenerate their fins, skin, and heart (Jopling, 2010; Sun et al., 2009). Jopling et al. (2010) demonstrated that [zebrafish] can fully regenerate its heart after amputation of up to 20 percent of the ventricle. Although very little is known about the average lifespan of green sea turtles, *Chelonia mydas* reach sexual maturity between 20 to 50 years of age (United States Fish and Wildlife Service, 2005).

2.2 Arriving at Reliable Models for Comparative Analysis

2.2.1 Performing Pairwise Sequence Alignment

Following group categorization, I deployed UGENE's pairwise sequence alignment (PSA) tool in order to identify similarity probabilities among my six protein sequences [length: 1,037]. Initially, I produced five pairwise sequence alignments by pairing each Klotho protein sequence against *Homo sapiens* [*type b*] and then grouping the results in numerical sequence [see below]. In this operation, Hirschberg algorithm was selected for optimal matching. A reliable algorithmic selection, Hirschberg increases efficiency by maximizing the sum of pairwise scores with quasi-gap penalties (Chao & Zhang, 2008). The Hirschberg algorithm can be derived from the Needleman-Wunsch algorithm by observing the following (Hirschberg, 1975):

if $(Z,W) = NW$[score] (X,Y) is the optimal alignment of (X,Y), and $X = X^l + X^r$ is an arbitrary partition of X, there exists a partition $Y^l + Y^r$ of Y such that NW[score] $(X,Y) = NW$[score] $(X^l,Y^l) + NW$[score] (X^r,Y^r)

When calculating the highest number of consistent patterns in a local alignment, gap penalty scores are often disregarded (Lassmann & Sonnhammer, 2005). However, because I am interested in obtaining a global alignment rather than calculating common subsequences, I applied a gap open penalty score of ten; including a gap extension penalty of two. The resulting PSA provided a first glimpse into similarity probabilities. As noted below, my numerically grouped PSA outputs infer high degree homology between three of the six sequences and two of six sequences; now categorized as group *type a* and group *type b*. PSA validates sequence placement among *type a* organisms and *type b* organisms.

Table 1. Similarity probabilities among *type a* and *type b* (PSA)

Group	Identification	Similarity appx.	FASTA description
type *b*	Homo sapiens	100%	>gi|2618596|dbj|BAA23382.1
type *b*	Mus musculus	85%	>gi|2618594|dbj|BAA23381.1
type *b*	Mesocricetus auratus	85%	>gi|524966649|ref|XP_005083219.1
type *a*	Heterocephalus glaber	54%	>gi|351702474|gb|EHB05393.1
type *a*	Chelonia mydas	52%	>gi|465963328|gb|EMP30051.1
type *a*	Danio rerio	50%	>gi|528492619|ref|XP_690797.4

2.2.2 Performing Multiple Sequence Alignment

A series of multiple sequence alignment (MSA) operations were then performed in preparation for phylogenetic analysis. In contrast to pairwise alignment, multiple sequence alignment can reveal subtle similarities among *groups* of proteins (Chao & Zhang, 2008). Here, I selected Kalign for multiple sequence alignment; an accurate and fast MSA algorithm (Lassmann & Sonnhammer, 2005). Kalign is an extension of Wu-Manber approximate pattern-matching algorithm, which is based on Levenshtein distances. This strategy enables Kalign to estimate sequence distances faster and more accurately than other popular iterative methods. Comparisons done by Lassmann and Sonnhammer (2005) show that Kalign is about 10 times faster than ClustalW and, depending on the alignment size, up to 50 times faster than other iterative methods; Kalign also delivers better overall resolution (Lassmann & Sonnhammer, 2005).

Kalign is noted for producing optimal execution times, and this procedure would require minimal computational resources. First, I initiated UGENE's multiple sequence alignment tool by importing and processing the six protein sequences in FASTA format. Due to parameter setting sensitivity in protein data types, Kalign for MSA gap penalty scores were modified slightly during successive intervals until an optimal global alignment was obtained. Each interval resulted in a 3,092 base-pair alignment, followed by a phylogenetic diagram. These operations were then repeated in successive fashion, upon conducting protein to nucleotide conversions.

2.3 Reverse KLOTHO Protein-DNA Translation for Phylogenetic Reconstruction

Generally speaking, protein sequences are intolerant of change in evolutionary context. Over the span of evolutionary time, protein sequences undergo selective constraints for protein function and protein structure, and these are conserved over much longer periods than individual codons (Martin & Palumbi, 1993). The most direct evolutionary changes to protein occur at the amino and carboxyl termini in the form of domain insertions, repetitions, and deletions (Marsh & Teichmann, 2010). Likewise, we must also consider the possibility that convergent evolution can occur to produce apparent similarity between proteins that are evolutionarily unrelated, but perform similar functions and have similar structures (Bastien, 2008). Thus, multiple substitutions at a single DNA base more accurately reflect mutational history (Martin & Palumbi, 1993).

The challenges in utilizing protein sequences to infer divergence events led me to consider a secondary option to protein-protein comparison in the context of phylogenetic reconstruction. Perhaps it would be necessary to cross-check the original diagram(s) produced by protein-protein comparison with a nucleotide derivative. Using my custom-based API translator (Protein to DNA Bio Translator), I reverse translated each sequence from its original data type, namely protein, into a workable [and theoretical] dataset made exclusively of KL-derivative nucleotide sequences (DNA).

Following conversion, the six DNA sequences were imported to UGENE's bioinformatics software, and upon file import and MSA execution, an alternate radial phylogenetic diagram was generated that would allow me to cross-confirm the original(s) obtained by protein-protein comparison. I implemented PHYLIP neighbor-joining method coupled with distance matrix model F84 on the 3,092 base-pair alignment; this procedure would also require additional bootstrapping compilers to help evaluate the strength of the nodes. Lastly, it is worth noting that the resulting phylogenetic tree(s) do not assume an evolutionary clock; it is in effect an unrooted tree.

The PHYLIP neighbor-joining method is capable of generating highly probable diagrams in scenarios involving low degrees of variance, regardless of dataset size. An accurate and statically consistent polynomial-time algorithm, PHYLIP neighbor-joining does not assume that all lineages evolve at the same rate (as proteins evolve at different rates), and it constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join [where a set of n taxa requires $n - 3$ iterations; each step is repeated by $(n - 1) \times (n - 1)$] (Felsenstein, 1981). For illustration purposes, the following formulas demonstrate a standard neighbor-joining Q-matrix algorithm:

$$Q(i,j) = (n - 2)\, d(i,j) - \sum \{n, k = 1\}\, d(i,k) - \sum \{n, k = 1\}\, d(j,k) \quad (1)$$

Pair to node (distances):

$$(f,u) = \tfrac{1}{2}\, d(f,g) + \tfrac{1}{2}(n - 2)\, [\, \sum \{n, k = 1\}\, d(f,k) - \sum \{n, k = 1\}\, d(g,k)\,] \quad (2)$$

Taxa to node (distances):

$$d(u,k) = \tfrac{1}{2}\, [\, d(f,k) + d(g,k) - d(f,g)\,] \quad (3)$$

*2.4 Analyzing α-*Amino Acid *Compositions*

Given its mass and length, the molecular weight measurements of protein structures are fundamentally important to its biochemical characterization and function. Moreover, the α-amino acid composition of each protein structure may contribute to the overall quality of a protein. If certain α-amino acids are optimal for protein structure, natural selection should have acted over evolutionary time to increase the frequency of these α-amino acids (Anthis et al., 2013). As noted by Mannakee and Gutenkunst (2012), catalytic domains in protein evolve faster, while non-catalytic domains in protein evolve more slowly. This may also suggest that networks typically evolve under stabilizing selection (Mannakee & Gutenkunst, 2012).

In addition to a phylogenetic reconstruction, this investigation puts a strong emphasis on discerning meaningful patterns, or sequence motifs, in lower and higher frequency α-amino acid compositions. By comparing the results of composition frequency percentages among group *type a* and group *type b*, I hope to identify any potentially significant molecular markers in evolutionary conserved chemical properties of Klotho that may have increased or decreased over evolutionary time, within a particular group and across a wider evolutionary spectrum. This procedure incorporates a few methods and techniques, such as α-amino acid residue calculations for determining molecular weight and frequency percentage.

A number of highly efficient web applications are well suited for α-amino acid composition analysis. In this phase of my investigation, I use Composition/Molecular Weight Calculation tool (University of Delaware, 2014) for obtaining the sum ratio of α-amino acid residues and approximate residue charge, and Protein Calculator (Anthis et al., 2013) to determine atomic compositions. I then apply the resulting figures and datasets in comparative analysis in order to evaluate the overall α-amino acid distributions.

The sum ratios in α-amino acid counts vary slightly due to irregularities in protein sequence length. This would also explain discrepancies in PSA similarity approximations. Amino acid counts are quantified by the total number of residues in an individual protein sequence. These figures are later arranged and disbursed as to determine frequency percentage within a protein sequence arrangement. The total estimate in molecular weight per arrangement is also deciphered by accounting for the total sum of individual residue weights. Lastly, my molecular weight formula integrates the following two variables: (x) the individual sum residue weight(s); (y) subtracted by the molecular weight of H_2O (18.01528) (University of Delaware, 2014).

Molecular weight = sum of individual residues weights - water molecular weight ′ (number of residues - 1)

a) Homo sapiens ($C_{5337}H_{8024}N_{1410}O_{1452}S_{30}$): 116131.78
b) Mesocricetus auratus ($C_{5357}H_{8070}N_{1434}O_{1449}S_{24}$): 116514.11
c) Mus musculus ($C_{5338}H_{8043}N_{1431}O_{1454}S_{28}$): 116424.91
d) Danio rerio ($C_{5178}H_{7752}N_{1380}O_{1437}S_{30}$): 113287.91
e) Chelonia mydas ($C_{3896}H_{5817}N_{979}O_{1045}S_{23}$): 83826.69
f) Heterocephalus glaber ($C_{3528}H_{5347}N_{919}O_{973}S_{21}$): 76876.45

3. Results

3.1 Reconstructing a Phylogeny Based on Klotho Gene Variants

The resulting base-pair alignments yielded nearly identical outputs. At first glance, *type a* organisms and *type b* organisms are each placed together within their respective groups – with four mammalian candidates on one end of the diagram(s), and the remaining non-mammalian candidates to the other [see Figure 1]. When cross-checked with morphology, the corresponding nodes reside in correct order of taxonomy on both diagrams. On one end, three of six organisms belonging to *order rodentia* coincide in proper placement according to scientific classification; while my fourth *mammalian* species (Homo sapiens) falls within a fairly close proximity; followed by the remaining candidates.

The phylogenies illustrated in this study depict gene families within paralogy regions consistent with the early evolution of vertebrates. Because similar variations of KL-derivative proteins occur in all species of *Chordata* – ranging from fish, reptiles to mammals – we may safely infer that, in fact, Klotho family proteins belong to ancient

paralogy lineages. While having the same basic molecular functions, Klotho protein variants may have undergone only tiny degrees of modification throughout evolutionary history (originating at the nucleotide level). As the next section will demonstrate, these tiny variations are further observed among my candidate sequences. Alas, the current MSA models help further verify group categorization, and the extent of evolutionary relatedness among group *type a* and group *type b* candidates, including node length and probable time scales in decimal format, are detailed below.

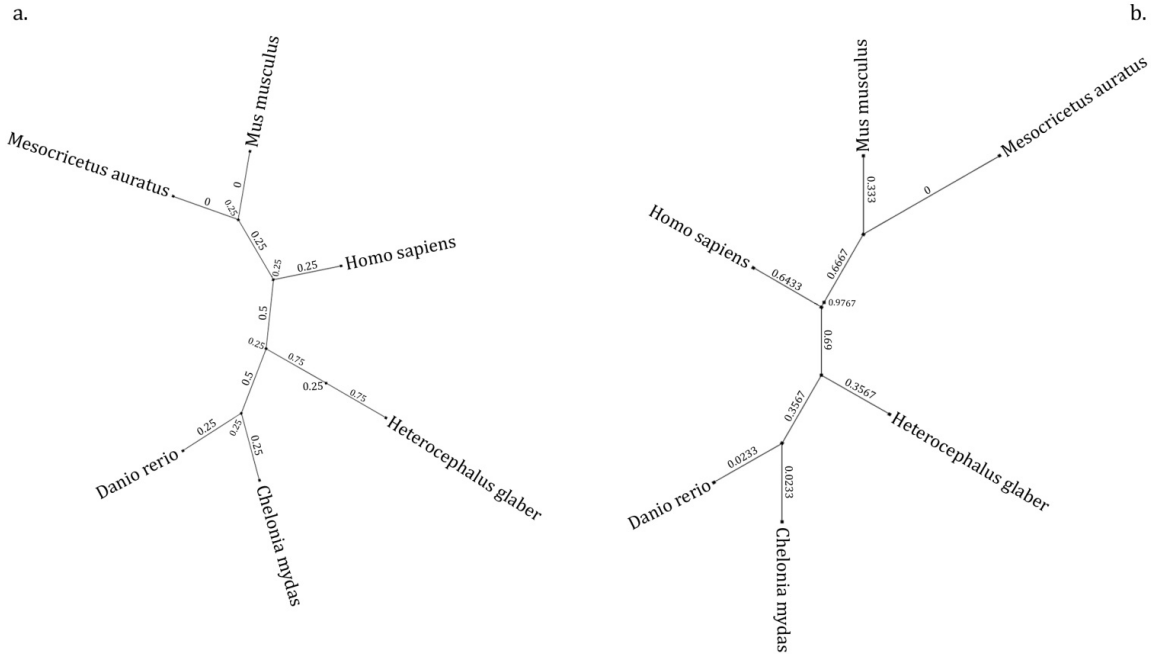

Figure 1. Radial phylogenetic trees. a) Protein-protein output (Klotho). b) Reverse protein-DNA output (Klotho)

3.2 Minimal Degrees of Variation in Evolutionary Conserved Chemical Properties of Klotho

The chemical properties of the α-amino acids of proteins determine the biological activity of the protein, and it conveys a vast array of chemical versatility (University of Arizona, 2014). Evolution itself exerts pressure to preserve α-amino acid residues that bear invaluable functional roles. According to Dokholyan et al. (2002), proteins may retain significantly critical chemical properties (or molecular signatures) that contribute, in lower or higher frequency values, to protein stability via their interaction with other α-amino acid residues. To this end, we might expect to observe a large percentage of evolutionary conserved α-amino acids across entire protein families, with only minimal degrees of variation occurring among wide ranging taxonomical groups.

Indeed, in the context of evolutionary conservation, only three of twenty α-amino acid**s** in this study exceed disparity ratios that average more than 0.093 percent. Comparative analyses of α-amino acid compositions highlight only three strongly distinct molecular signatures indicative of each group. As Figure 2 and Figure 3 illustrate, a correlation between elevated levels of L-Lysine coupled with low quantities of L-Alanine and L-Arginine are featured in group *type a* [as compared to group *type b*]; where the average disparity ratio is ±1.8 percent (L-Lysine), ±1.6 percent (L-Alanine), and ±1.34 percent (L-Arginine). A max disparity ratio of >3.1 percent occurs in L-Lysine molecules [*type a*]. Increased levels of L-Lysine are significant in terms of optimal structure and fold, and suggest that natural selection may have acted to increase the frequency of this molecule.

It is worthwhile to note that L-Alanine and L-Arginine are manufactured internally via biosynthetic pathways; whereas L-Lysine must be ingested as lysine or lysine-containing proteins (plants synthesize L-Lysine from aspartic acid) (Beals, 1999). While the implications of dietary L-Lysine are inconsequential to a study involving techniques in comparative sequence analysis, various other studies suggest that increased dietary L-Lysine plays a role in calcium absorption; a function that is also inadvertently linked to KLOTHO (Tan et al., 2012). As noted in the sections above, "Klotho stimulates calcium reabsorption in the distal convoluted tubule by deglycosylating and stabilizing the epithelial calcium channel TRPV5 on the surface of cellular membrane (Robinson, 2012)." This, and proficiencies in tissue regeneration and other immunity-related health benefits are also linked to increased levels of dietary L-Lysine (Falco, 1995).

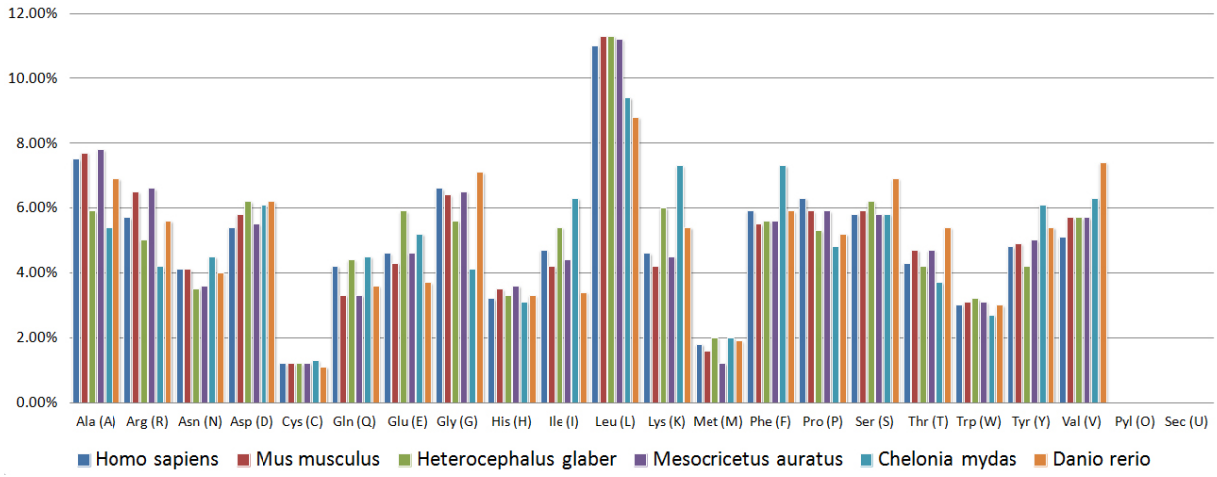

Figure 2. Amino acid composition analysis (six sequences)

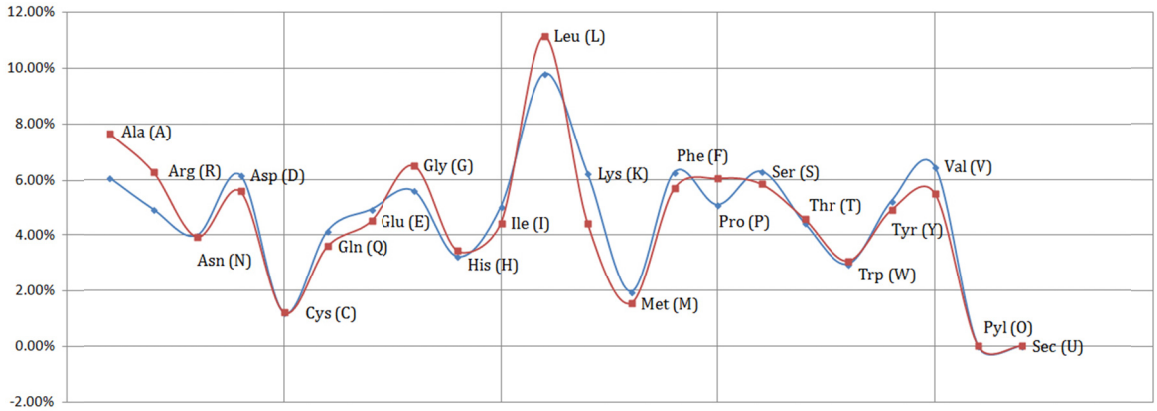

Figure 3. Amino acid composition analysis (*type a* highlighted in blue; *type b* highlighted in red)

4. Discussion

Proteins are essential building blocks of living cells; and life itself can be viewed as resulting substantially from the chemical activity of proteins (Lupas, 2014). Systematic approaches directed toward the evolution of proteins suggest that protein interactions were possibly the first forms of self-reproducing life on earth. As such, the chemical properties of each protein have been developed over billions of years. The α-amino acid composition of proteins contributes to the overall quality of a protein. If certain α-amino acids are optimal for protein structure, natural selection should have acted over evolutionary time to increase the frequency of those α-amino acids (Anthis et al., 2013).

As previously noted, the chemical properties of proteins are also conserved over much longer periods than individual codons; and frequency patterns in Klotho gene variants across wide ranging taxonomy confirm it. Yet, conflicting distribution ratios among L-Lysine, L-Alanine, and L-Arginine [in group *type a* versus group *type b*] is indicative of natural selection acting upon frequency distribution. Could this have consequences in terms of protein function that contribute toward Klotho-associated anti-aging mechanisms?

In fully examining the data presented in this study, it is tempting for me to draw premature [and definitive] conclusions based upon the preliminary appearance of a modest correlation; but correlation does not necessarily imply causation. Instead of over-speculating on functional cause-and-effect and probable implications thereof, I would simply stress the primary scope of this paper is not to presume on solutions, but rather, on associations and identification. From the perspective of comparative sequence analysis, I have outlined a set of significantly distinct biomarkers in molecular variation patterns of KLOTHO across wide ranging taxonomy; and, in the broadest sense, I demonstrated homology and relatedness based on protein variants of Klotho that is consistent with the early evolution of vertebrates. There is still considerable work to do to improve our understanding of age-related

mechanisms and the protein functions that play a decisive role. And while a great deal about protein evolution remains unresolved, much can be learned about protein evolution by trying to reconstruct ancient paralogy lineages via comparative analysis of modern proteins. With this in mind, I believe that my results may help provide meaningful reference toward a fully systematic investigation.

References

United States Fish and Wildlife Service. (2005). *Green sea turtle (Chelonia mydas)*. North Florida Field Office.

American Federation for Aging Research. (2006). *Theories of aging -- Why and how do we age?*

Florida State University. (2007). Chemists Kill Cancer Cells With Light-activated Molecules. *ScienceDaily*. Retrieved from http://www.sciencedaily.com/releases/2007/08/070808132019.htm

Minnesota State University. (2012). Molecular Weight Estimations and SDS PAGE.

University of Arizona. (2014). *The Chemistry of Amino Acids*.

Agnisola, C., & Femiano, S. (2006). Studies on routine metabolism in adult zebrafish, Danio rerio. *Acta Physiol., 187*.

Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical Science,* 364-376.

Andreollo, N. A., Santos, E. F. D., Araújo, M. R., & Lopes, L. R. (2012). Rat's age versus human's age: what is the relationship? ABCD. *Arquivos Brasileiros de Cirurgia Digestiva (São Paulo), 25*(1), 49-51. http://dx.doi.org/10.1590/S0102-67202012000100011

Anthis, N. J., & Clore, G. M. (2013). Sequence-specific determination of protein and peptide concentrations by asorbance at 205 nm. *Protein Science, 22*, 851-858. http://dx.doi.org/10.1002/pro.2253

Bastien, O, Aude, J. C., Roy, S., & Marechal, E. (2004). Fundamentals of massive automatic pairwise alignments of protein sequences: theoretical significance of Z-value statistics. *Bioinformatics, 20*(4), 534–537. http://dx.doi.org/10.1093/bioinformatics/btg440

Beals, M., Gross, L., & S. Harrell, S. (1999). *Amino Acid Frequency*. University of Tennessee. Retrieved from http://www.tiem.utk.edu/~gross/bioed/webmodules/α-amino acid.htm

Bouchard, C., Tremblay, A., Nadeau, A., Despres, J. P., Theriault, G., Boulay, M. R., ... & Fournier, G. (1989). Genetic effect in resting and exercise metabolic rates. *Metabolism, 38*(4), 364-370. http://dx.doi.org/10.1016/0026-0495(89)90126-1

Bryan, J. K. (1980). Aspartate family and branched-chain amino acids. In B. J. Miflin (Ed.), *The Biochemistry of Plants* (Vol. 5, pp. 403-452). New York: Academic Press.

Buffenstein, R. (2008). Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *Journal of Comparative Physiology B, 178*(4), 439-445. http://dx.doi.org/10.1007/s00360-007-0237-5

Chao, K. M., & Zhang, L. (2008). *Sequence comparison: theory and methods* (Vol. 7). Springer.

Chen, C., Sander, J. E., & Dale, N. M. (2003). The effect of dietary lysine deficiency on the immune response to Newcastle disease vaccination in chickens. *Avian diseases, 47*(4), 1346-1351. http://dx.doi.org/10.1637/7008

Chen, L., Bush, D. R. (1997). LHT1, a lysine- and histidine-specific amino acid transporter in Arabidopsis. *Plant Physiol., 115*, 1127-1134. http://dx.doi.org/10.1104/pp.115.3.1127

Composition/Molecular Weight Calculation tool. (ND). University of Delaware. Retrieved from http://pir.georgetown.edu/pirwww/search/comp_mw.shtml

Csiszar, A., Labinskyy, N., Orosz, Z., Xiangmin, Z., Buffenstein, R., & Ungvari, Z. (2007). Vascular aging in the longest-living rodent, the naked mole rat. *American Journal of Physiology-Heart and Circulatory Physiology, 293*(2), H919-H927. http://dx.doi.org/10.1152/ajpheart.01287.2006

DALY, T. J. M., WILLIAMS, L. A., & BUFFENSTEIN, R. (1997). Catecholaminergic innervation of interscapular brown adipose tissue in the naked mole‐rat (Heterocephalus glaber). *Journal of anatomy, 190*(3), 321-326. http://dx.doi.org/10.1046/j.1469-7580.1997.19030321.x.

Dena, B. (2014). Life Extension Factor Klotho Enhances Cognition, Cell Reports, http://dx.doi.org/10.1016/j.celrep.2014.03.076.

Dokholyan, N. V., Mirny, L. A., & Shakhnovich, E. I. (2002). Understanding conserved amino acids in proteins. *Physica A: Statistical Mechanics and its Applications, 314*(1), 600-606. http://dx.doi.org/10.1016/S0378-4371(02)01079-8

Enstipp, M. R., Ciccione, S., Gineste, B., Milbergue, M., Ballorain, K., Ropert-Coudert, Y., ... & Georges, J. Y. (2011). Energy expenditure of freely swimming adult green turtles (Chelonia mydas) and its link with body acceleration. *The Journal of experimental biology, 214*(23), 4010-4020. http://dx.doi.org/10.1242/jeb.062943

Falco, S. C., Guida, T., Locke, M., Mauvais, J., Sanders, C., Ward, R. T., & Webber, P. (1995). Transgenic canola and soybean seeds with increased lysine. *Biotechnology (N.Y.), 13*, 577-582. http://dx.doi.org/10.1038/nbt0695-577

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution, 17*(6), 368–376. http://dx.doi.org/10.1007/BF01734359.

Galili, G. (1995). Regulation of lysine and threonine synthesis. *Plant Cell, 7*, 899-906. http://dx.doi.org/10.1105/tpc.7.7.899

Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., ... & Gill, R. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature, 428*(6982), 493-521. http://dx.doi.org/10.1038/nature02426

Griffith, R. S., Norins, A. L., & Kagan, C. (1978). A multicentered study of lysine therapy in herpes simplex infection. *Dermatology, 156*(5), 257-267. http://dx.doi.org/10.1159/000250926

Hirschberg, D. S. (1975). A linear space algorithm for computing maximal common subsequences. *Communications of the ACM, 18*(6), 341–343. http://dx.doi.org/10.1145/360825.360861.

IUPAC, I. (1984). Nomenclature and symbolism for amino acids and peptides. *Eur J Biochem, 138*, 9-37.

Jopling, C., Sleep, E., Raya, M., Martí, M., Raya, A., & Belmonte, J. C. I. (2010). Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature, 464*(7288), 606-609. http://dx.doi.org/10.1038/nature08899

Jumpertz, R., Hanson, R. L., Sievers, M. L., Bennett, P. H., Nelson, R. G., & Krakoff, J. (2011). Higher energy expenditure in humans predicts natural mortality. *The Journal of Clinical Endocrinology & Metabolism, 96*(6), E972-E976. http://dx.doi.org/10.1210/jc.2010-2944

Kim, B. (2008). Thyroid hormone as a determinant of energy expenditure and the basal metabolic rate. *Thyroid, 18*(2), 141-144. http://dx.doi.org/10.1089/thy.2007.0266

Kurosu, H., Yamamoto, M., Clark, J. D., Pastor, J. V., Nandi, A., Gurnani, P., … Kuro-o, M. (2005). Suppression of aging in mice by the hormone Klotho. *Science, 309*(5742), 1829-1833. http://dx.doi.org/10.1126/science.1112766

Lassmann, T., & Sonnhammer, E. L. (2005). Kalign—An Accurate and Fast Multiple Sequence Alignment Algorithm. *BMC Bioinformatics, 6*, 298. http://dx.doi.org/10.1186/1471-2105-6-298

Little, A. G., & Seebacher, F. (2014). Thyroid hormone regulates cardiac performance during cold acclimation in zebrafish (Danio rerio). The Journal of experimental biology, 217(5), 718-725. http://dx.doi.org/10.1242/jeb.096602

Mannakee, B., & Gutenkunst, R. (2012). *Influence of Protein Network Dynamics on Protein Evolution*. University of Arizona.

Marsh, J. A., & Teichmann, S. A. (2010). How do proteins gain new domains. *Genome Biol, 11*(7), 126. http://dx.doi.org/10.1186/gb-2010-11-7-126

Martin, A. P., & Palumbi, S. R. (1993). Body size, metabolic rate, generation time, and the molecular clock. *Proceedings of the National Academy of Sciences, 90*(9), 4087-4091. http://dx.doi.org/10.1073/pnas.90.9.4087

Mueller, P., & Diamond, J. (2001). Metabolic rate and environmental productivity: well-provisioned animals evolved to run and idle fast. *Proceedings of the National Academy of Sciences, 98*(22), 12550-12554. http://dx.doi.org/10.1073/pnas.221456698

NHLBI Communications Office. (2004). *Scientists Compare Rat Genome With Human, Mouse -- Analysis Yields New Insights into Medical Model, Evolutionary Process*. National Institutes of Health.

Pérez, V. I., Buffenstein, R., Masamsetti, V., Leonard, S., Salmon, A. B., Mele, J., ... & Chaudhuri, A. (2009). Protein stability and resistance to oxidative stress are determinants of longevity in the longest-living rodent, the naked mole-rat. *Proceedings of the National Academy of Sciences, 106*(9), 3059-3064. http://dx.doi.org/10.1073/pnas.0809620106

Robinson, M. (2012). *In response to: Which is more informative: a phylogenetic tree based on alignment of*

protein α-amino acid sequences or one based on the corresponding DNA sequences? Los Alamos National Laboratory. Retrieved from Research Gate.

Rozing, M. P., Westendorp, R. G., de Craen, A. J., Frölich, M., Heijmans, B. T., Beekman, M., ... & van Heemst, D. (2009). Low serum free triiodothyronine levels mark familial longevity: the Leiden Longevity Study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, glp200.

Speakman, J. R., Selman, C., McLaren, J. S., & Harper, E. J. (2002). Living fast, dying when? The link between aging and energetics. *The Journal of nutrition, 132*(6), 1583S-1597S.

Sun, P., Zhang, Y., Yu, F., Parks, E., Lyman, A., Wu, Q., ... & Hsiai, T. K. (2009). Micro-electrocardiograms to study post-ventricular amputation of zebrafish heart. *Annals of biomedical engineering, 37*(5), 890-901. http://dx.doi.org/10.1007/s10439-009-9668-3

Tan, T. C., Rahman, R., Jaber-Hijazi, F., Felix, D. A., Chen, C., Louis, E. J., & Aboobaker, A. (2012). Telomere maintenance and telomerase activity are differentially regulated in asexual and sexual worms. *Proceedings of the National Academy of Sciences, 109*(11), 4209-4214. http://dx.doi.org/10.1073/pnas.1118885109

Torres, P. U., Prie, D., Molina-Bletry, V., Beck, L., Silve, C., & Friedlander, G. (2007). Klotho: an antiaging protein involved in mineral and vitamin D metabolism. *Kidney international, 71*(8), 730-737. http://dx.doi.org/10.1038/sj.ki.5002163

Vogt, G., Etzold, T., & Argos, P. (1995). An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited. *Journal of molecular biology, 249*(4), 816-831. http://dx.doi.org/10.1006/jmbi.1995.0340

Vogt, G., Etzold, T., & Argos, P. (1995). An assessment of amino acid exchange matrices in aligning protein sequences: the twilight zone revisited. *Journal of molecular biology, 249*(4), 816-831. http://dx.doi.org/10.1006/jmbi.1995.0340

Warwicker, J., Charonis, S., & Curtis, R. A. (2013). Lysine and Arginine Content of Proteins: Computational Analysis Suggests a New Tool for Solubility Design. *Molecular pharmaceutics, 11*(1), 294-303. http://dx.doi.org/10.1021/mp4004749

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

Gene Expression Patterns in Functionally Different Cochlear Compartments of the Newborn Rat

Johann Gross[1], Heidi Olze[1] & Birgit Mazurek[1]

[1] Molecular Biology Research Laboratory, Department of Otorhinolaryngology, Charité Universitätsmedizin Berlin, Campus Charité Mitte, Berlin, Germany

Correspondence: Johann Gross, Molecular Biology Research Laboratory, Department of Otorhinolaryngology, Charité Universitätsmedizin Berlin, Campus Charité Mitte, Berlin, Germany. E-mail: johann.gross@charite.de; johann.gross@arcor.de

Received: December 17, 2014 Accepted: January 3, 2014 Online Published: January 12, 2015

doi:10.5539/jmbr.v5n1p20 URL: http://dx.doi.org/10.5539/jmbr.v5n1p20

Abstract

In an experimental model of organotypic cultures of the stria vascularis (SV), the organ of Corti (OC) and the modiolus (MOD), we compared the expression levels and injury/hypoxia induced response of 36 genes associated with the cells´ energy-producing and energy-consuming processes, using the microarray technique. A decrease of expression was observed for most of the voltage-dependent K^+- and Ca^{++}- channels as an effective mechanism to lower energetic demands. We identified two gene networks of transcripts that are differentially expressed across the three regions. One cluster is associated with the transcription factor hypoxia-inducing factor (*Hif-1a*) and the second one with the caspase and calpain cell death genes *Casp3, Capn1, Capn2* and *Capns1*. The *Hif-1a* gene subset consists of genes belonging to the glucose metabolism (glucose transporter *Slc2a1*, glycolytic enzymes *Gapdh*, *Hk1* and *Eno2*), the Na^+/K^+ homeostasis (ATPase *Atp1a1*) and the glutamate pathway (NMDA receptor associated protein 1 *Grina*, glutamate transporter *Slc1a1*, *Slc1a3*). The *Slc2a1*, *Gapdh*, *Hk1*, *Slc1a3*, *Grina* and *Atp1a1* transcripts are also members of the cell death subset indicating a role they have to play in the differential regional cell death rates. The newly identified genes *Grina* and calnexin (*Canx*) may play specific and yet unknown roles in regulating cell death induced by injury and hypoxia in the inner ear. We assume that the differential regional response occurs on the basis of endogenous gene regulatory mechanisms and may be important to maintaining the cochlea's function following damage from trauma and hypoxia.

Keywords: cell death, gene expression, hypoxia, injury, inner ear, microarray

1. Introduction

The cochlea consists of three main complex structures, each serving a specific function: the organ of Corti (OC), the modiolus (MOD) and the stria vascularis (SV). The OC with its inner and outer hair cells transforms the mechanical signal into an electrical one via hair cell depolarization and signal amplification. The specific function of the SV is to produce and maintain the ionic composition of the endolymph, a very specific fluid with high concentrations of potassium and low concentrations of sodium. The MOD, the conically shaped central axis in the cochlea, contains the spiral ganglion neurons (SGNs). These bipolar neuronal cells transmit the electrical signals from the hair cells to the cochlear nuclei in the brainstem.

Organotypic cultures of the SV, the OC and the MOD were used to experimentally study the differential gene expression of these regions to injury stress and hypoxia (Gross et al., 2007). In freshly prepared tissue, about 2-10 % of all nuclei were found to be stained by propidium iodide, indicating cell damage during preparation of the cochlear tissue. After 24h in culture, the number of necrotic cells in the MOD region increased from 8 % to 25-35 %), whereas the number of such cells remained unchanged in the OC and SV regions (Gross et al., 2008). Gene expression markers indicate that two basic pathogenetic mechanisms are involved in this experimental model: mechanically induced inflammation and hypoxia. The expression of genes involved in apoptosis and necrosis (*Casp3, Capn1, Capn2* and *Capns1*), reactive oxygen species metabolism (*Sod3, Nos2*), inflammation (*Ccl20*) as well as selected transcription factors (*Hif-1a*, *Jun,* and *Bmyc*) have to play a key role in the differential regional response (Mazurek et al., 2011; Gross, Olze, & Mazurek, 2014).

Other pathways implicated in injury and hypoxia include processes involved in energy production and regulation of ion homeostasis which could become critical for cell survival and regeneration (Michiels, 2004). The energy balance of a

cell under physiological and pathological conditions depends on its ATP production and consumption. The energy demand of the inner ear is high and can be compared to that of brain tissue. To meet the energy demand under conditions of inflammation and hypoxia, anaerobic ATP supply is triggered (Pasteur effect; Boutilier & St-Pierre, 2000). ATP production via glycolysis is associated with increased glucose transport and consumption (Frezza et al., 2011).

Maintaining ion homeostasis to allow de- and repolarization of the cells belongs to the processes that require high amounts of energy (Rolfe & Brown, 1997; Buttgereit & Brand, 1995). In the cochlea, the maintenance of ion homeostasis, K^+-cycling and its role in the endocochlear potential are coupled to Na,K-ATPase (Wangemann, 2002). K^+-cycling is particularly important, as in response to the stimulated stereocilia, endolymphatic K^+ flows into the sensory hair cells via the apical transduction channel and is released from the hair cells into the perilymph via basolateral K^+ channels. K^+ may be taken up by fibrocytes in the spiral ligament and transported from cell to cell via gap junctions into strial intermediate cells which secrete it to the endolymph.

Calcium entry and the maintenance of the multiple segregated transduction pathways is controlled by a combination of calcium channels, Ca^{++}-ATPases and buffering mechanisms. Elevated levels of intracellular calcium under hypoxia or injury stress are the result of a massive influx of extracellular calcium through activated channels or the release of calcium from intracellular stores like the endoplasmatic reticulum or the mitochondria (Brini & Carafoli, 2009; Lang, Vallon, Knipper, & Wangemann, 2007). Main routes for calcium influx are voltage-gated Ca^{++} channels, purinergic receptors and ionotropic glutamate receptors (Martinez-Sanchez et al., 2004). For active efflux of intracellular Ca^{++}, the main routes have found to be its export via the plasma membrane calcium ATPase (PMCA) and the transport via the Na^+/Ca^{++} exchanger. Another mechanism to decrease cytosolic Ca^{++} concentrations is uptake into intracellular stores via the sarco- and endo-plasmic reticulum calcium ATPase (SERCA). The calcium buffering mechanisms consist of several Ca^{++} binding proteins.

The aim of the present study is to analyze the basal and injury-induced expression of genes associated with the energy-producing and energy-consuming processes, i.e., of genes associated with the glucose metabolism, the regulation of Na^+-/K^+- and Ca^{++}-homeostasis, including the glutamate pathway. A microarray study was used as the guide that directed us to selecting a total of 36 genes associated with energy production and energy consumption, including the glutamate pathway (Mazurek et al., 2006). The use of the neurobiological array RN-U34 offers the possibility of identifying several transcripts in the inner ear that had not been described previously.

2. Materials and Methods

2.1 Explant Cultures

The cochleae from 3 to 5-day old Wistar rats were dissected into OC, MOD and SV (Sobkowicz, Loftus, & Slapnick, 1993). Details of the preparation of the fragments, the culture conditions and the testing of the viability of the explants were reported previously (Gross et al., 2007). Briefly, the fragments were incubated in four-well tissue culture dishes in Dulbecco`s Modified Eagle Medium/F12Nutrient (1:1) Mixtures (Gibco, Karlsruhe, Germany) supplemented with 10 % fetal bovine serum. Fragments of one ear were kept in culture under normoxic conditions, fragments of the second ear were exposed to moderate hypoxia (oxygen partial pressure inside the culture medium was 10-20 mm Hg) for 5 h, starting three hours after plating. The number of dead cells was determined in freshly prepared tissue (controls) and after 24 h in culture using the live/dead viability test by propidium iodide (PI) and calcein AM staining (Gross et al., 2008).

2.2 cDNA Microarray Analysis

The cDNA microarray analysis was carried out using the Affymetrix Rat Neurobiology U34 Array (RN-U34; Affymetrix, Santa Clara, USA). The complete data sets from this study have been deposited to the Gene Expression Omnibus (GEO) database according to the MIAME standard and can be accessed by ID GSE5446. Each of the total RNA samples of the MOD, OC and SV used in the microarray study originated from 6 animals. Altogether, 16 RNA preparations arising from three independent series were analyzed within one year: four samples from freshly prepared tissue (OC1, OC2, MOD, SV) and 12 experimental samples from cultures of OC, MOD and SV under normoxic (n = 2) and hypoxic conditions (n = 2). Further details of the RNA isolation and quantification and the cDNA microarray analysis were previously reported (Gross et al., 2007).

2.3 Statistical Analysis

Intensity of expression was classified on the basis of the histogram of normalized log2 signals and resulted in a normal distribution (data not shown). Values at or above the 75th percentile of the cumulative intensities are considered to be high level expression (> log2 = 12.55, = 6000 relative units; bold in Tables 1-3) and values below the 25th percentile to be low level expression (< log2 = 10.43, = 1380 relative units; italics in Tables 1-3); values between them are considered to be moderate (normal typeface in Tables 1-3). In this study, gene expression was

not found to differ significantly between normoxic and hypoxic environments, neither for the numbers of PI-stained nuclei nor for the expression of HIF-1a mRNA or for that of other genes, with the data having been combined to result in four samples per region. Obviously, the hypoxia conditions we used were too mild to induce specific expression changes. Overlapping gene expression patterns induced by hypoxia and mechanical injury may also contribute to this observation. Three features of the gene expression are presented: (i) The absolute expression levels, classified as low, moderate or high using the log2 data. (ii) The fold change of the expression level was calculated as the ratio between the expression intensity of the 24h cultures and the expression intensity of freshly prepared tissue. The mean coefficient of the variation was 11.8 ± 7.6 % (n = 36) for signal intensity and 17.6 ± 8.1 % (n = 108; Tables 1-3) for the expression change. To test the significance levels of the fold changes we used the paired t-test or the Wilcoxon paired test (Tables 1-3). (iii) We used the Pearson´s correlation analysis with the Bonferroni post hoc test to identify co-expression changes among selected transcripts across the three regions.

3. Results

3.1 Glucose Transporter and Glycolytic Enzymes

The chip comprises the transcripts of glucose transporters *Slc2a1* (Glut1) and *Slc2a3* (Glut3) and the glycolytic enzymes *Gapdh* (glyceraldehyde-3-phosphate dehydrogenase)*, Hk1* (hexokinase 1) and *Eno2* (neuron-specific enolase; Table 1). As expected, the *Slc2a1* and the *Gapdh* transcripts show high expression levels in all regions, whereas *Hk1* and *Eno2* belong to the subset of genes with low or moderate expression levels. In culture, with the exception of *Eno2* in the MOD region, all transcripts increase significantly across the three regions.

Table 1. Expression of genes associated with glucose transport and glycolysis in the organotypic cultures of the modiolus, the organ of Corti and the stria vascularis

Gene	Expression			Fold change			Name/function
	MOD	OC	SV	MOD	OC	SV	
Slc2a1 (S68135)*[1]	**12109**	**8328**	**19903**	2.4	7.9[#1]	3.6	Glucose transp. 1 (Glut1)
Slc2a3 (D13962)*[2]	1764	1542	*1132*	1.3	2.1	3.4	Glucose transp. 3 (Glut3)
Gapdh (X02231.1)*[3]	**15383**	**24992**	**24879**	7.2	5.4	5.0	Gap-Dehydrogenase
Hk1 (J04526.1)*[4]	*1012*	1556	1866	2.2[#2]	1.6	1.3	Hexokinase 1
Eno2 (X07729)	5069	4818	1235	0.9	1.9[#3]	5.0[#4]	Enolase 2, gamma

Note. Expression intensity (relative units, RU) was categorized on the basis of the histogram of the normalized log2 signals (see materials and methods). Fold change (Wilcoxon paired test, n = 12): *[1]T(12) = 0.00, $p = 0.002$; *[2]T(12) = 0.00, $p = 0.002$; *[3]T(12) = 2.00, $p = 0.004$; *[4]T(12) = 2.00, $p = 0.004$. Glut - glucose transporter, Gap-glyceraldehyde-3-phosphate. [#]Significance of expression changes (paired t-test, n = 4): [#1]p<0.000 vs MOD, p<0.006 vs SV, [#2]p<0.02 vs SV, [#3]p<0.004 vs MOD, [#4]p<0.000 vs MOD, 0.006 vs OC.

3.2 Na^+/K^+-Homeostasis

Several hundred genes known to encode ion channel proteins are involved in the regulation of Na^+/K^+-homeostasis (Gabashvili, Sokolowski, Morton, & Giersch, 2007). The chip identified the transcript for one sodium and six potassium channels (Table 2), most of them previously not identified in the inner ear. These channels show low to moderate basal expression levels and do not change (*Scn3a, Kcnk1, Kcnq3* and *Kcns3*) or decrease (*Pias3, Alg10* and *Kcnh2*). The basal expression levels and the fold changes in culture are similar in all regions.

Sodium-potassium ATPase is of importance not only for the neuronal resting potential but also for re-establishing the ion homeostasis in the inner ear following injury and hypoxia (Johar, Priya, & Wong-Riley, 2012; Wangemann, 2002). The chip comprises two isoforms of the alpha subunit and two isoforms of the beta subunit (Table 2). The basal expression levels of the various isoforms are unevenly distributed in the three regions and, as expected, relatively high in the OC and SV regions. The most remarkable findings in the present work are the increase of the *Atp1a1* subunit and the decrease of the *Atp1a3, Atp1b2* and the *Atp1b3* subunits.

3.3 Calcium Homeostasis

Calcium entry and the maintenance of the multiple segregated transduction pathways is controlled by a combination of voltage-dependent calcium channels, purinergic and ionotropic glutamate receptors, the Ca^{++}-ATPases and several buffering mechanisms (Table 3).

3.3.1 Calcium Channels

The chip contains information for several voltage-dependent Ca^{++}-channels which respond to cell membrane potential changes and have a role to play in changes of the local intracellular Ca^{++} homeostasis and the formation of the nerve impulse (*Cacna1d* /D38101, *Cacna1g/*AF027984, *Cacna2d1/*M8662, *Cacnb3/*M88751). These transcripts show low to moderate basal expression levels and decrease in culture 0.4-0.7-fold (data not shown). The expression of the voltage-dependent calcium channel *Cacn1c* and of the purinergic channel *P2rx* remained unchanged in all regions.

Table 2. Expression of genes associated with Na^+/K^+ -transport in the organotypic cultures of the modiolus, the organ of Corti and the stria vascularis

Gene	Expression			Fold change			Name/function
	MOD	OC	SV	MOD	OC	SV	
Scn3a (Y00766)	*1206*	*1159*	*897*	0.8	0.8	0.7	VG Na^+channel, type III, alpha
Kcnk1 (AF022819)	*780*	3514	2542	1.2	1.2	1.2	K^+-channel, SF K, M1 (Twik)
Kcnq3 (AF091247)	*1280*	*1333*	*1084*	1.1	0.8	0.8	K^+-VGCh, SF KQT-like, M3
Kcns3 (Y17607)	*404*	2211	1433	1.0	0.7	1.0	K^+- VGCh, SF S, M3
Pias3 (AF032872)*[1]	4281	3549	3688	0.5	0.7	0.5	Binding protein to Kv-channels
Alg10 (U78090)*[2]	3299	3180	2446	0.6	0.8	0.5	Regulatory component
Kcnh2 (U75210)*[3]	4063	3846	2847	0.6	0.6	0.7	K^+-VGCh, SF H, M2 (ERG1)
Atp1a1 (M74494)*[4]	1498	**6001**	**15083**	**8.3**	**4.1**	**2.9**	ATPase, Na^+/K^+ transp., alpha1
Atp1a3 (M28648)*[5]	2475	5905	*1207*	0.2	0.4	0.3	ATPase, Na^+/K^+ transp., alpha3
Atp1b2 (J04629)*[6]	3914	**6888**	**2628**	0.5	0.4	0.3	ATPase, Na^+/K^+ transp., beta2
Atp1b3 (D84450)*[7]	**16726**	**8315**	**14474**	0.7	0.9	0.8	ATPase, Na^+/K^+ transp., beta3

Note. Expression see legend to Table 1. Abbr.: VGCh-Voltage gated channel; SF-subfamily; M-Member; V-voltage; Pias3 - Protein inhibitor of activated STAT, 3; Alg10 - Asparagine-linked glycosylation 10**,** regulatory component of non-inactivating K^+ channels, voltage-gated K^+ channel binding protein, involved in modulating the expression of Kv2 channels; transp.- transporting. Fold change (Wilcoxon paired test, n = 12): $^{*1\text{-}7}T(12) = 0.00$, $p = 0.002$.

3.3.2 Glutamate Pathway

The following transcripts involved in regulating the activity of the glutamate pathway were detected by the chip: three sequences associated with the ionotropic glutamate receptors (*Grina, Grin2B*, and *Grik5*), the glutamate receptor interacting protein 2 (*Grip2*), two glutamate transporters (*Slc1a1* and *Slc1a3*) and the glutamate-ammonia ligase (*Glul*). These transcripts have not been previously characterized in the inner ear. *Grina* encodes for a glutamate-binding subunit of an NMDA receptor-associated complex protein (NMDARA1), also called glutamate-binding protein (Kumar, Tilakaratne, Johnson, Allen, & Michaelis, 1991; Nielsen et al., 2011). It is characterized by a high basal expression level in all regions; in culture, its expression level remained unchanged in the OC and the SV and tended to increase in the MOD region. *Grin2b* encodes for the ionotropic NMDA2B receptor (NR2B), *Grik5* for the ionotropic kainate 5 receptor and *Grip2* for the glutamate interacting protein 2, which binds and affects AMPA receptors. These transcripts are moderately expressed and decrease in culture in all regions (Table 3).

Slc1a1 and *Slc1a3*, two glutamate transporters, are members of the excitatory amino acid transporter (EAAT) family of high-affinity sodium-dependent glutamate carriers encoded by the genes of the SLC1 family. *Slc1a1* encodes for the solute carrier family 1 (the neuronal/epithelial high affinity glutamate transporter, also known as EAAT3, EAAC1; Chen, Kujawa, & Sewell, 2010b). *Slc1a1* shows a low basal expression, and in culture, its expression tends to increase in the SV. *Slc1a3* encodes for the solute carrier family 1 member 3 (glial high affinity glutamate transporter, also known as EAAT1, GLAST). This transporter shows a high basal expression level; in culture, its expression decreases clearly in all regions. *Glul* encodes for the glutamate-ammonia ligase (also known as glutamine synthetase) which converts glutamate to glutamine; this transcript shows a high basal expression level, and in culture, its expression tends to increase in the OC.

Table 3. Expression of genes associated with Ca^{++} homeostasis in the modiolus, the organ of Corti and the stria vascularis

Gene	Expression			Fold change			Name/function
	MOD	OC	SV	MOD	OC	SV	
Cacna1c (M59786)	2396	*1292*	1845	0.9	1.1	0.8	VCa-ch, L-type,a-1C
P2rx2 (AF020756)	4585	**10920**	**9326**	0.7	1.3	1.1	Purinerg. rec. P2X, ch2
P2rx4 (U47031)	2416	1882	2194	0.9	0.9	1.1	Purinerg. rec. P2X, ch4
Grina (S61973)	**10282**	**21127**	**9227**	1.7[#1]	0.8	0.9	GR-NMDA-ass.prot.1
Grin2b (U11419)*[1]	2864	2414	2292	0.8	0.7	0.7	GR-NMDA2B
Grip2 (AF090113)*[2]	2794	2400	2052	0.7	0.7	0.7	GR-interact.prot.2
Grik5 (Z11581)*[3]	2567	3917	3010	0.6	0.6	0.8	GR-kainate5
Slc1a1 (D63772)	1658	*1060*	*827*	0.8	1.1	2.4[#2]	Eaat3
Slc1a3 (S59158)*[4]	**12059**	**23828**	**9227**	0.2[#3]	0.6	0.5	Eaat1
Glul (M91652)*[5]	**19899**	**17553**	**36544**	1.3	1.8[#4]	1.1	Glul
Atp2b1 (L04739)	1538	1486	*1120*	0.8	0.7	0.9	Pmca1b
Atp2b2 (J03754)*[6]	*785*	2190	*716*	0.1	0.5	0.3	Pmca2
Calm2 (M17069)*[7]	**37541**	**37770**	**33732**	0.5	0.7	0.7	Calmodulin 2
Calm3 (X14265)	**6740**	5556	5074	0.8	1.2	1.0	Calmodulin 3
Canx (L18889)*[8]	5700	**7439**	5467	2.2	2.2	2.6	Calnexin
Ppp3ca (D90035)*[9]	**6015**	**6021**	5400	0.9	0.8	0.7	Calcineurin

Note. Expression see legend to Table 1. Abbr.: VCa-ch – Voltage dependent calcium channel; Purinerg. rec.- purinergic receptor, ligand-ion channel; GR-NMDA-ass.prot.1 - glutamate receptor, N-methyl D-aspartate-associated protein 1; NMDA2B - glutamate receptor, ionotrop, N-methyl D-aspartate 2B; GR-interact.prot.2- GR-interacting protein 2; Eaat3 -neuronal/epithelial high affinity glutamate transporter, system Xag, member 1; Eaat1 - glial high affinity glutamate transporter, member 3; Glul - glutamate-ammonia ligase. Pmca – plasma membrane calcium ATPase. Fold change (Wilcoxon paired test, n = 12): $^{*1,2}T(12) = 0.00$, $p = 0.002$; $^{*3}T(12) = 1.00$, $p = 0.003$; $^{*4\text{-}9}T(12) = 0.00$, $p = 0.002$; $^{\#1}p<0.005$ vs OC and 0.001 vs SV; $^{\#2}p<0.022$ vs MOD; $^{\#3}p<0.001$vs OC; $^{\#4}p<0.02$ vs SV (n = 4).

3.3.3 Calcium ATPases and Calcium Binding Proteins

Ca^{++}-ATPases contribute largely to re-establishing Ca^{++}- ion homeostasis after unregulated Ca^{++}- influx, which is an energy demanding process (Buttgereit & Brand, 1995). The chip comprises two transcripts for plasma membrane calcium ATPase (PMCA; *Atp2b1*, *Atp2b2;* Table 3) and two for smooth endoplasmic reticulum calcium ATPase (SERCA; *Atp2a2*, *Atp2a3*). *Atp2b1* encodes for PMCA1, an enzyme with a housekeeping function. *Atp2b2* encodes for PMCA2, an enzyme with special functions in maintaining Ca^{2+} homeostasis in hair cells (Brini & Carafoli, 2009). Unlike PMCAs, SERCAs accumulate Ca^{++} into vesicles of the endoplasmic reticulum at the expense of ATP hydrolysis. The SERCA transcripts *Atp2a2* (J04739) and *Atp2a3* (M30581) show moderate expression levels and decrease significantly in all regions (0.5 - 0.6 fold, data not shown).

The present array data showed high to moderate expression levels for the Ca^{++}-binding proteins calmodulin (CaM) *Calm2 and Calm3,* for calnexin (*Canx*) and for calcineurin (*Ppp3ca*; Table 3). In culture, *Calm2* and *Ppp3ca* expression decreased in all regions, whereas *Canx* increased. *Calm3* remained unchanged. These features are in line with the finding that calcium-binding proteins constitute a high portion of the total cellular protein in all mammalian cells and are involved in protecting from calcium overload.

3.4 Co-Expression Analysis

Hypoxia inducible factor (HIF) is a key transcription factor regulating adaptation to hypoxia and tissue injury and it plays an important part in cell survival (Semenza, 2001). To characterize possible associations between the expression changes of *Hif-1a* and the transcripts analyzed in this study, we correlated the expression changes across the three regions. We observed that *Hif-1a* expression has correlations to nine transcripts associated with

the metabolism of glucose (*Slc2a1, Slc2a3, Gapdh, Hk1* and *Eno2)*, Na^+/K^+ homeostasis (*Atp1a1*) and the glutamate pathway (*Slc1a1, Slc1a3, Grina;* Figure 1A). A more detailed analysis of the data sets shows that the significance between two genes includes a different regional response (Figure 1B-E). The glucose metabolism associated genes belong to the classical target genes of HIF-1alpha (Greijer & van der Wall, 2004; Marin-Hernandez, Gallardo-Perez, Ralph, Rodriguez-Enriquez, & Moreno-Sanchez, 2009; Yu et al., 2012). Remarkably, the higher *Hif-1a* expression in the MOD region is associated with a relative decrease of expression levels of *Slc1a1/a3* and a relative increase of the *Grina* transcript.

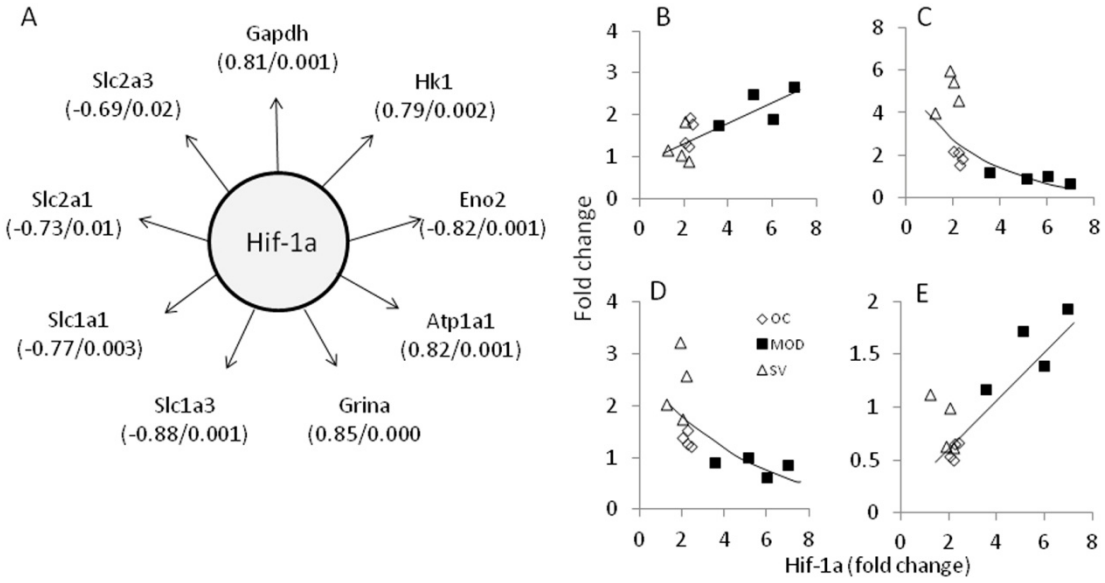

Figure 1. Relationship between *Hif-1a* expression and transcripts belonging to the glucose metabolism, the Na^+, K^+ homeostasis and the glutamate pathway

Note. (A) Diagram illustrating the cluster of transcripts associated with *Hif-1a* expression. Numbers indicate correlation coefficients *r* and significance levels *p*, n = 12. Marginal significance was observed for *Slc2a1* and *Slc2a3*. (B-E) Examples of scatter plots illustrating the correlations between *Hif-1a* and selected transcripts of Figure 1A. B - *Hk1*; C – *Eno2;* D - *Slc1a1;* E - *Grina*. The best fit curve to the *Eno2* and *Slc1a1* data is an exponential function.

To identify co-expression changes across the three regions in relation to cell death, we correlated the changes of these transcripts with molecules known to be mediators in apoptotic and necrotic cell death (Gross, Olze, & Mazurek, 2014). The present study shows that the *Slc2a1, Slc1a3, Grina, Atp1a1, Gapdh* and *Hk1* transcripts correlate closely with the cell death subunits (Figure 2A). These correlations are based on different responses in the MOD region compared to OC and SV (Figure 2B-E).

4. Discussion

4.1 Up-Regulated Transcripts

The up- and down-regulation of transcript levels appear plausible and efficient in terms of energy expenditure of the underlying processes. In the present study, we observe an up-regulation of transcripts involved in energy production and protective mechanisms. It is known that an increased mRNA synthesis which may contribute to an increased transcript levels is very energy-demanding (Simpson, Carruthers, & Vannucci, 2007). The parallel increase of the glucose transporter transcripts *Slc2a1* and *Slc2a3*, in particular in the OC and SV regions, and the up-regulation of *Gapdh* and *Hk1* may primarily contribute to increasing energy production (Edamatsu, Kondo, & Ando, 2011; Marin-Hernandez, Gallardo-Perez, Ralph, Rodriguez-Enriquez, & Moreno-Sanchez, 2009). The differential expression of *Hk1* and *Eno2* may have additional functions. The mitochondrial-bound isoform HK1 may interact with the membrane permeability transition (MPT) pore through the voltage-dependent anion channel (VDAC) which inhibits the cytochrome c release induced by the pro-apoptotic proteins Bax and Bid (Azoulay-Zohar, Israelson, Abu-Hamad, & Shoshan-Barmatz, 2004; Marin-Hernandez, Gallardo-Perez, Ralph,

Rodriguez-Enriquez, & Moreno-Sanchez, 2009). *Eno2* encodes for the neuron-specific enolase (NSE) which is the gamma-gamma enolase isoenzyme. The unchanged expression level in the MOD and the increase of *Eno2* in the OC and the SV may have a role to play in the cells' adaptation to stress (Yan et al., 2011). The expression increase of *Atp1a1* in all regions underlies the important functional role of Na,K-ATPase for cell survival, as studied using specific inhibitors (Johar, Priya, & Wong-Riley, 2012; Fu, Ding, Jiang, & Salvi, 2012). The calcium-binding protein *Canx* is important as it assumes the role of a chaperone in order to transport newly synthesized proteins from the endoplasmatic reticulum to the outer cellular membrane (Zuppini et al., 2002). Previous work showed that *Grina* is a member of the transmembrane BAX inhibitor motif (TMBIM3) known as an anti-apoptotic protein that controls apoptosis through the modulation of ER calcium homeostasis (Rojas-Rivera et al., 2012). *Grina's* high basal expression level and its increased expression in the MOD may well be involved in cell protection (Goswami et al., 2012). What is of interest here is the specific increase of *Slc1a1* (glutamate transporter) in the SV and of *Glul* (glutamate-ammonia ligase) in the OC. There is *in vitro* evidence of a polarized brain-to-blood transport of glutamate by endothelial cells co-cultured with astrocytes (Helms, Madelung, Waagepetersen, Nielsen, & Brodin, 2012). GLUL converts glutamate to glutamine and may contribute to the elimination of toxic glutamate (Takumi et al., 1997).

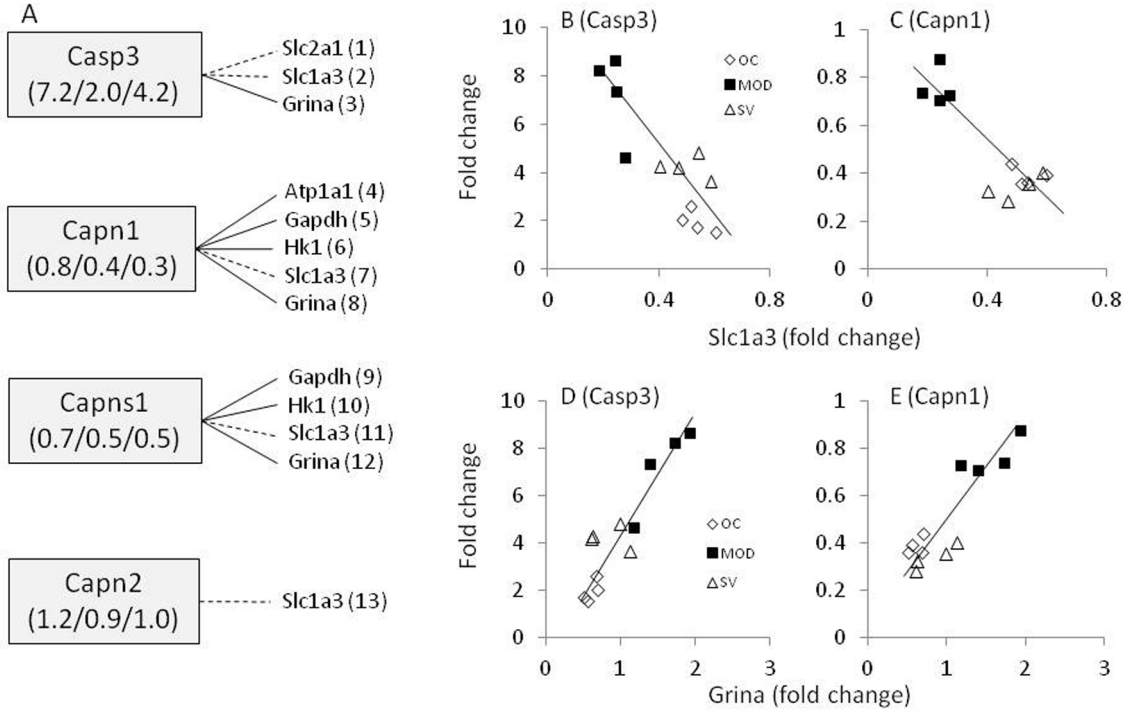

Figure 2. Members of the cell death cluster

Note. (A) Diagram illustrating the cluster of transcripts involved in apoptosis and necrosis. Lines 1-13 indicate significant correlations with correlation coefficients in the range $r = 0.78 – 0.92$ and significance levels in the range $p < 0.000 – 0.003$. Broken lines indicate negative correlations. Numbers within the rectangle indicate the fold change in MOD/OC/SV (Gross, Olze, & Mazurek, 2014). (B-E) Examples of scatter plots illustrating the correlations between transcripts of the glutamate system and cell death transcripts. Correlation coefficients and significance levels (r/p): B,- 0.87/0.000; C, -0.85/0.000; D, 0.92/0.000; E, 0.90/0.000.

4.2 Down-Regulated Transcripts

An important way of adapting to the energetic deficit following tissue injury and hypoxia is to decrease major ATP consuming functions. Several subsets of transcripts are characterized by a clear expression decrease, among them the K^+ and Ca^{++} - channels, the Atpase subunits *Atp1a3*, *Atp1b2*, the two SERCA transcripts, three glutamate receptors and the glutamate transporter subunit *Slc1a3*. Certain hypoxia-tolerant lower vertebrate species resort to decreasing ion channels as a mechanism to lower their energetic demands (Boutilier & St-Pierre, 2000). Because macromolecule turnover and ion-motive ATPases are major ATP consumers it comes as no surprise in the present

study that the mRNA levels of ion channels are found to be down-regulated. Suppression of ion channel densities probably associated with lower cell membrane permeability decreases the energetic costs of maintaining electrochemical gradients (so-called 'channel arrest'; Hochachka, Buck, Doll, & Land, 1996). Activities that are essential to the maintenance of life should be able to function at lower energy charge values (Atkinson, according to Buttgereit & Brand, 1995; Wieser & Krumschnabel, 2001). Whereas the changes of most of the transcripts may be advantageous for cell survival, the down-regulation of the glial high affinity glutamate transporter *Slc1a3* appears to exert a rather damaging effect, in particular in the MOD region, because this strong expression decrease of *Slc1a3* may contribute to elevated extracellular glutamate concentrations and cell damage in the MOD region (Bianchi, Bardelli, Chiu, & Bussolati, 2014; Gegelashvili & Schousboe, 1997).

4.3 Transcripts Without Significant Changes

Several K^+ channels, the purinergic ion channels, the voltage-dependent calcium channel *Cacn1c* and calmodulin *Calm3* belong to the subset of transcripts without significant changes. The unchanged expression of these channels may allow speculations to be made about the importance of these particular transcripts to cell survival (Rolfe & Brown, 1997). The *Cacn1c* subunit is part of the Cav1.2 channel that plays an important role in synaptic-activity-dependent gene expression and may be important for regenerative processes. With the exception of the *Cacnb3* data, no data are available for these channels in the inner ear (Kuhn et al., 2009). The high expression levels of the calcium-binding proteins *Calm3* and of the purinergic ion channel *P2rx2* in the OC and the SV made us assume that the corresponding proteins have a role to play in cell survival or cell repair. Calcineurin plays an important role in refilling Ca^{2+} stores of the endoplasmatic reticulum and maintaining optimum conditions for protein processing and folding (Bollo et al., 2010).

4.4 Hif-1a Associated Genes

Observations to the extent that the energy balance, glucose uptake by the cells, ion homeostasis and glutamatergic neurotransmission are interlinked, support the assumption of a membership of these genes in a functional gene network, with *HIF-1a* being an important regulator of efficient adaptation across the three regions (Greene & Greenamyre, 1996; Rodriguez-Rodriguez, Almeida, & Bolanos, 2013). For example, glutamate transport depends on the ATPase to remove Na^+ from the cytoplasm, and the activity of Na,K-ATPase is dependent on ATP production (Casey, Pakay, Guppy, & Arthur, 2002). Our data suggest that the differential expression changes of *Atp1a1*, the two glutamate transporters and of *Grina* optimize the adaption process of the glutamate system during repair and regeneration.

The co-expression changes between *Hif-1a* and several genes on the transcript levels could be explained by unique features observed for HIF-1 expression in recent years. Several authors showed a crucial role for Hif-1a mRNA turnover to exist in HIF-1 signaling, and a regulatory role of mRNA turnover as a modulator of HIF-1a function, independent of the oxygen tension (Fahling et al., 2012; Schodel, Mole, & Ratcliffe, 2013; Eltzschig & Carmeliet, 2011). Little is known about the factors that regulate expression changes of *Grina* and of the glutamate transporters. Endothelins, a family of peptides up-regulated in the injured brain, negatively regulate glial glutamate transporter expression (Rozyczka, Figiel, & Engele, 2004). Studies in astrocytes indicate that the suppression of *Slc1a3* is Ca^{++}-dependent (Liu, Yang, & Tzeng, 2008). By inhibiting the down-regulation of *Slc1a3* and buffering the glutamate homeostasis, taurine protects retinal cells *in vitro* under hypoxic conditions (Chen et al., 2010a).

Other than for these genes, we observed no significant correlations between *Hif-1a* and Ca^{++}-ATPases. The expression and activity of proteins involved in Ca^{2+} regulation are subject to the autoregulatory principle, which means that they are regulated by the Ca^{2+} signal itself (Brini & Carafoli, 2009). Other investigations have shown that the level of expression of proteins associated with Ca^{++} homeostasis is regulated by transcription (recently reviewed by Ritchie, Zhou, & Soboloff, 2011).

4.5 Cell Death Associated Genes

The present work suggests that *Slc2a1, Slc1a3, Grina, Atp1a1, Gapdh* and *Hk1* may be involved in the differential cell death rate. The type of cell death immediately after the preparation of the cultures can be categorized as accidental necrotic cell death (ANCD), whereas the cell death measured 24h after the damaging event corresponds most probably to secondary necrotic cell death (SNCD) or late apoptotic cell death (Krysko et al., 2011). The observed genetic responses may be important in terms of secondary cell-death prevention. Preventing the glutamate transporter *Slc1a3* from strongly decreasing in the MOD region or the increase of *Grina* can be assumed to be of special importance to the survival of the SGNs following injury and hypoxia. Up to now, it is unclear which factors show the highest regulatory strength to induce cell death or to maintain survival.

4.6 Conclusions

This study documents previously undescribed genetic features of the different compartments of the inner ear. Both gene clusters, the *Hif-1a* (Figure 1) and the cell death cluster (Figure 2), are based on the co-expression changes of genes across OC, MOD and SV and are the result of a differential, tissue-specific response to injury and hypoxia. The different response occurs on the basis of endogenous gene regulatory mechanisms developed in the course of evolution. The differential response appears important to maintain the cochlea's function following damage from environmental factors. For example, the decrease of ion channels in all regions or the differential response of *Hif-1a* and other transcription factors may be crucial for inner ear function.

The basal expression and the injury/hypoxia-induced patterns may contribute to the region-related difference in the cell death rates immediately after injury and after 24 h in culture. The newly identified genes *Grina* and *Canx* may play specific and yet unknown roles in regulating cell death induced by injury and hypoxia in the inner ear.

However, caution is indicated in the interpretation of these data for several reasons. First, the noise of microarray might influence some of the experimental results. However, the significant correlations between microarray data and quantitative RT-PCR values observed in several studies, justify our approach. Second, the data are not complete in the sense that only some members of pathways have been experimentally determined. Third, we are aware that the response of immature tissue is different from that of mature tissue. Nevertheless, responses are very similar to that in mature tissue (Kennedy, 2012). Fourth, the analysis quantified mRNA levels, but the data do not indicate whether subsequent proteins are generated and where they are located. Beyond the up and down-regulation of transcript expression, the question arises of whether such changes are functional. However, many similarities in the response of the corresponding proteins, even in *in vivo* studies, lead to the conclusion that these observations are far more important than only for the present model.

Acknowledgements

We would like to thank the University Hospital Charité for support. It gives us great pleasure to thank Johannes Wendt for his generous help in critically reading and correcting this article.

References

Azoulay-Zohar, H., Israelson, A., Abu-Hamad, S., & Shoshan-Barmatz, V. (2004). In self-defence: hexokinase promotes voltage-dependent anion channel closure and prevents mitochondria-mediated apoptotic cell death. *Biochem. J., 377*, 347-355. http://dx.doi.org/10.1042/BJ20031465

Bianchi, M. G., Bardelli, D., Chiu, M., & Bussolati, O. (2014). Changes in the expression of the glutamate transporter EAAT3/EAAC1 in health and disease. *Cell Mol. Life Sci., 71*, 2001-2015. http://dx.doi.org/10.1007/s00018-013-1484-0

Bollo, M., Paredes, R. M., Holstein, D., Zheleznova, N., Camacho, P., & Lechleiter, J. D. (2010). Calcineurin interacts with PERK and dephosphorylates calnexin to relieve ER stress in mammals and frogs. *PLoS. ONE., 5*, e11925. http://dx.doi.org/10.1371/journal.pone.0011925

Boutilier, R. G., & St-Pierre, J. (2000). Surviving hypoxia without really dying. *Comp Biochem. Physiol A Mol. Integr. Physiol., 126*, 481-490. http://dx.doi.org/10.1016/S1095-6433(00)00234-8

Brini, M., & Carafoli, E. (2009). Calcium pumps in health and disease. *Physiol Rev., 89*, 1341-1378. http://dx.doi.org/10.1152/physrev.00032.2008

Buttgereit, F., & Brand, M. D. (1995). A hierarchy of ATP-consuming processes in mammalian cells. *Biochem. J., 312*(Pt 1), 163-167.

Casey, T. M., Pakay, J. L., Guppy, M., & Arthur, P. G. (2002). Hypoxia causes downregulation of protein and RNA synthesis in noncontracting Mammalian cardiomyocytes. *Circ. Res., 90*, 777-783. http://dx.doi.org/10.1161/01.RES.0000015592.95986.03

Chen, F., Mi, M., Zhang, Q., Wei, N., Chen, K., Xu, H., ... Chang, H. (2010a). Taurine buffers glutamate homeostasis in retinal cells in vitro under hypoxic conditions. *Ophthalmic Res., 44*, 105-112. http://dx.doi.org/10.1159/000312818

Chen, Z., Kujawa, S. G., & Sewell, W. F. (2010b). Functional roles of high-affinity glutamate transporters in cochlear afferent synaptic transmission in the mouse. *J. Neurophysiol., 103*, 2581-2586. http://dx.doi.org/10.1152/jn.00018.2010

Edamatsu, M., Kondo, Y., & Ando, M. (2011). Multiple expression of glucose transporters in the lateral wall of the cochlear duct studied by quantitative real-time PCR assay. *Neurosci. Lett., 490*, 72-77. ttp://dx.doi.org/10.1016/j.neulet.2010.12.029

Eltzschig, H. K., & Carmeliet, P. (2011). Hypoxia and inflammation. *N. Engl. J Med., 364*, 656-665. http://dx.doi.org/10.1056/NEJMra0910283

Fahling, M., Persson, A. B., Klinger, B., Benko, E., Steege, A., Kasim, M., ... Mrowka, R. (2012). Multilevel regulation of HIF-1 signaling by TTP. *Mol. Biol. Cell., 23*, 4129-4141. http://dx.doi.org/10.1091/mbc.E11-11-0949

Frezza, C., Zheng, L., Tennant, D. A., Papkovsky, D. B., Hedley, B. A., Kalna, G., … Gottlieb, E. (2011). Metabolic profiling of hypoxic cells revealed a catabolic signature required for cell survival. *PLoS. ONE., 6*, e24411. http://dx.doi.org/10.1371/journal.pone.0024411

Fu, Y., Ding, D., Jiang, H., & Salvi, R. (2012). Ouabain-induced cochlear degeneration in rat. *Neurotox. Res., 22*, 158-169. http://dx.doi.org/10.1007/s12640-012-9320-0

Gabashvili, I. S., Sokolowski, B. H., Morton, C. C., & Giersch, A. B. (2007). Ion channel gene expression in the inner ear. *J Assoc. Res. Otolaryngol., 8*, 305-328. http://dx.doi.org/10.1007/s10162-007-0082-y

Gegelashvili, G., & Schousboe, A. (1997). High affinity glutamate transporters: regulation of expression and activity. *Mol Pharmacol., 52*, 6-15.

Goswami, D. B., Jernigan, C. S., Chandran, A., Iyo, A. H., May, W. L., Austin, M. C., … Karolewicz, B. (2012). Gene expression analysis of novel genes in the prefrontal cortex of major depressive disorder subjects. Prog. Neuropsychopharmacol. *Biol. Psychiatry., 43C*, 126-133. http://dx.doi.org/10.1016/j.pnpbp.2012.12.010

Greene, J. G., & Greenamyre, J. T. (1996). Bioenergetics and glutamate excitotoxicity. *Prog. Neurobiol., 48*, 613-634. http://dx.doi.org/10.1016/0301-0082(96)00006-8

Greijer, A. E., & van der Wall, W. E. (2004). The role of hypoxia inducible factor 1 (HIF-1) in hypoxia induced apoptosis. *J. Clin. Pathol., 57*, 1009-1014. http://dx.doi.org/10.1136/jcp.2003.015032

Gross, J., Machulik, A., Amarjargal, N., Moller, R., Ungethum, U., Kuban, R. J., … Mazurek, B. (2007). Expression of apoptosis related genes in the organ of Corti, modiolus and stria vascularis of newborn rats. *Brain Res., 1162*, 56-68. http://dx.doi.org/10.1016/j.brainres.2007.05.061

Gross, J., Machulik, A., Moller, R., Fuchs, J., Amarjargal, N., Ungethuem, U., … Mazurek, B. (2008). mRNA expression of members of the IGF system in the organ of Corti,the modiolus and the stria vascularis of newborn rats. *Growth Factors., 26*, 180-191. http://dx.doi.org/10.1080/08977190802194317

Gross, J., Olze, H., & Mazurek, B. (2014). Differential Expression of Transcription Factors and Inflammation-, ROS-, and Cell Death-Related Genes in Organotypic Cultures in the Modiolus, the Organ of Corti and the Stria Vascularis of Newborn Rats. *Cell Mol. Neurobiol., 34*, 523-538. http://dx.doi.org/10.1007/s10571-014-0036-y

Helms, H. C., Madelung, R., Waagepetersen, H. S., Nielsen, C. U., & Brodin, B. (2012). In vitro evidence for the brain glutamate efflux hypothesis: brain endothelial cells cocultured with astrocytes display a polarized brain-to-blood transport of glutamate. *Glia., 60*, 882-893. http://dx.doi.org/10.1002/glia.22321

Hochachka, P. W., Buck, L. T., Doll, C. J., & Land, S. C. (1996). Unifying theory of hypoxia tolerance: molecular/metabolic defense and rescue mechanisms for surviving oxygen lack. *Proc. Natl. Acad. Sci. U. S. A., 93*, 9493-9498. http://dx.doi.org/10.1073/pnas.93.18.9493

Johar, K., Priya, A., & Wong-Riley, M. T. (2012). Regulation of Na+/K+-ATPase by nuclear respiratory factor 1: Implication in the tight coupling of neuronal activity, energy generation, and energy consumption. *J Biol. Chem., 287*, 40381-40390. http://dx.doi.org/10.1074/jbc.M112.414573

Kennedy, H. J. (2012). New developments in understanding the mechanisms and function of spontaneous electrical activity in the developing mammalian auditory system. *J Assoc. Res. Otolaryngol., 13*, 437-445. http://dx.doi.org/10.1007/s10162-012-0325-4

Kuhn, S., Knirsch, M., Ruttiger, L., Kasperek, S., Winter, H., Freichel, M., ... Engel, J. (2009). Ba2+ currents in inner and outer hair cells of mice lacking the voltage-dependent Ca2+ channel subunits beta3 or beta4. *Channels (Austin.)., 3*, 366-376. http://dx.doi.org/10.4161/chan.3.5.9774

Kumar, K. N., Tilakaratne, N., Johnson, P. S., Allen, A. E., & Michaelis, E. K. (1991). Cloning of cDNA for the glutamate-binding subunit of an NMDA receptor complex. *Nature., 354*, 70-73. http://dx.doi.org/10.1038/354070a0

Krysko, D. V., Agostinis, P., Krysko, O., Garg, A. D., Bachert, C., Lambrecht, B. N., & Vandenabeele, P. (2011). Emerging role of damage-associated molecular patterns derived from mitochondria in inflammation. *Trends Immunol., 32*, 157-164. http://dx.doi.org/10.1016/j.it.2011.01.005

Lang, F., Vallon, V., Knipper, M., & Wangemann, P. (2007). Functional significance of channels and transporters expressed in the inner ear and kidney. *Am J Physiol Cell Physiol., 293*, C1187-C1208. http://dx.doi.org/10.1152/ajpcell.00024.2007

Liu, Y. P., Yang, C. S., & Tzeng, S. F. (2008). Inhibitory regulation of glutamate aspartate transporter (GLAST) expression in astrocytes by cadmium-induced calcium influx. *J Neurochem., 105*, 137-150. http://dx.doi.org/10.1111/j.1471-4159.2007.05118.x

Marin-Hernandez, A., Gallardo-Perez, J. C., Ralph, S. J., Rodriguez-Enriquez, S., & Moreno-Sanchez, R. (2009). HIF-1alpha modulates energy metabolism in cancer cells by inducing over-expression of specific glycolytic isoforms. *Mini. Rev. Med. Chem., 9*, 1084-1101. http://dx.doi.org/10.2174/138955709788922610

Martinez-Sanchez, M., Striggow, F., Schroder, U. H., Kahlert, S., Reymann, K. G., & Reiser, G. (2004). Na(+) and Ca(2+) homeostasis pathways, cell death and protection after oxygen-glucose-deprivation in organotypic hippocampal slice cultures. *Neuroscience., 128*, 729-740. http://dx.doi.org/10.1016/j.neuroscience.2004.06.074

Mazurek, B., Amarjargal, N., Haupt, H., Fuchs, J., Olze, H., Machulik, A., & Gross, J. (2011). Expression of genes implicated in oxidative stress in the cochlea of newborn rats. *Hear. Res., 277*, 54-60. http://dx.doi.org/10.1016/j.heares.2011.03.011

Mazurek, B., Machulik, A., Amarjargal, N., Kuban, R. J., Ungethuem, U., Fuchs, J., ... Gross, J. (2006). Gene expression of organ of Corti (OC), modiolus (MOD) and stria vascularis (SV) of newborn rats. Gene Expression Omnibus (GEO) website (http://www.ncbi.nlm.nih.gov/geo/). ID GSE5446.

Michiels, C. (2004). Physiological and pathological responses to hypoxia. *Am J Pathol., 164*, 1875-1882. http://dx.doi.org/10.1016/S0002-9440(10)63747-9

Nielsen, J. A., Chambers, M. A., Romm, E., Lee, L. Y., Berndt, J. A., & Hudson, L. D. (2011). Mouse transmembrane BAX inhibitor motif 3 (Tmbim3) encodes a 38 kDa transmembrane protein expressed in the central nervous system. *Mol. Cell Biochem., 357*, 73-81. http://dx.doi.org/10.1007/s11010-011-0877-3

Ritchie, M. F., Zhou, Y., & Soboloff, J. (2011). Transcriptional mechanisms regulating Ca(2+) homeostasis. *Cell Calcium., 49*, 314-321. http://dx.doi.org/10.1016/j.ceca.2010.10.001

Rodriguez-Rodriguez, P., Almeida, A., & Bolanos, J. P. (2013). Brain energy metabolism in glutamate-receptor activation and excitotoxicity: role for APC/C-Cdh1 in the balance glycolysis/pentose phosphate pathway. Neurochem. *Int., 62*, 750-756. http://dx.doi.org/10.1016/j.neuint.2013.02.005

Rojas-Rivera, D., Armisen, R., Colombo, A., Martinez, G., Eguiguren, A. L., Diaz, A., … Hetz, C. (2012). TMBIM3/GRINA is a novel unfolded protein response (UPR) target gene that controls apoptosis through the modulation of ER calcium homeostasis. *Cell Death. Differ., 19*, 1013-1026. http://dx.doi.org/10.1038/cdd.2011.189

Rolfe, D. F., & Brown, G. C. (1997). Cellular energy utilization and molecular origin of standard metabolic rate in mammals. *Physiol Rev., 77*, 731-758.

Rozyczka, J., Figiel, M., & Engele, J. (2004). Endothelins negatively regulate glial glutamate transporter expression. *Brain Pathol., 14*, 406-414. http://dx.doi.org/10.1111/j.1750-3639.2004.tb00084.x

Schodel, J., Mole, D. R., & Ratcliffe, P. J. (2013). Pan-genomic binding of hypoxia-inducible transcription factors. *Biol. Chem., 394*, 507-517. http://dx.doi.org/10.1515/hsz-2012-0351

Semenza, G. L. (2001). Hypoxia-inducible factor 1: control of oxygen homeostasis in health and disease. *Pediatr. Res., 49*, 614-617. http://dx.doi.org/10.1203/00006450-200105000-00002

Simpson, I. A., Carruthers, A., & Vannucci, S. J. (2007). Supply and demand in cerebral energy metabolism: the role of nutrient transporters. *J Cereb. Blood Flow Metab., 27*, 1766-1791. http://dx.doi.org/10.1038/sj.jcbfm.9600521

Sobkowicz, H. M., Loftus, J. M., & Slapnick, S. M. (1993). Tissue culture of the organ of Corti. *Acta Otolaryngol. Suppl (Stockholm)., 502*, 3-36.

Takumi, Y., Matsubara, A., Danbolt, N. C., Laake, J. H., Storm-Mathisen, J., Usami, S., ... Ottersen, O. P. (1997). Discrete cellular and subcellular localization of glutamine synthetase and the glutamate transporter GLAST in the rat vestibular end organ. *Neuroscience., 79*, 1137-1144. http://dx.doi.org/10.1016/S0306-4522(97)00025-0

Wangemann, P. (2002). K+ cycling and the endocochlear potential. *Hear. Res., 165*, 1-9. http://dx.doi.org/10.1016/S0378-5955(02)00279-4

Wieser, W., & Krumschnabel, G. (2001). Hierarchies of ATP-consuming processes: direct compared with indirect measurements, and comparative aspects. *Biochem. J., 355*, 389-395. http://dx.doi.org/10.1042/0264-6021:3550389

Yan, T., Skaftnesmo, K. O., Leiss, L., Sleire, L., Wang, J., Li, X., & Enger, P. O. (2011). Neuronal markers are expressed in human gliomas and NSE knockdown sensitizes glioblastoma cells to radiotherapy and temozolomide. *BMC. Cancer., 11*, 524. http://dx.doi.org/10.1186/1471-2407-11-524

Yu, J., Li, J., Zhang, S., Xu, X., Zheng, M., Jiang, G., & Li, F. (2012). IGF-1 induces hypoxia-inducible factor 1alpha-mediated GLUT3 expression through PI3K/Akt/mTOR dependent pathways in PC12 cells. *Brain Res., 1430*, 18-24. http://dx.doi.org/10.1016/j.brainres.2011.10.046

Zuppini, A., Groenendyk, J., Cormack, L. A., Shore, G., Opas, M., Bleackley, R. C., & Michalak, M. (2002). Calnexin deficiency and endoplasmic reticulum stress-induced apoptosis. *Biochemistry., 41*, 2850-2858. http://dx.doi.org/10.1021/bi015967+

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

αS1-Casein Lineage Assessed by RFLP in the Endangered Goat Breed "Retinta Extremeña"

José L. Fernández-García[1] & María P. Vivas Cedillo[1]

[1] Genetics and Animal Breeding, Veterinary Faculty, Universidad de Extremadura, 10071 Cáceres, Spain

Correspondence: José L. Fernández-García, Genetics and Animal Breeding, Veterinary Faculty, Universidad de Extremadura, 10071 Cáceres, Spain. Tel: 34-927-257-105. E-mail: pepelufe@unex.es

Received: December 7, 2014 Accepted: December 29, 2014 Online Published: March 3, 2015

doi:10.5539/jmbr.v5n1p32 URL: http://dx.doi.org/10.5539/jmbr.v5n1p32

Abstract

The Retinta Extremeña goat is a well-adapted breed to "Dehesa" environment. Traditionally their raw milk is used to make artisan cheese. However, crosses with specialized breeds are occurring since the eighties, this goat breed has been declared of special protection by the Spanish Agriculture Ministry (R.D. 1682/1997 and R.D. 229/2008). Genetic studies about casein variants have been mainly performed on Spanish goats of high milk yields because the caseins are a relevant fraction of milk. But recent studies claimed to study the caseins in all breeds, including threatened goat breeds to decide about its conservation value. This study was focused on the αS1-casein in the endangered "Retinta Extremeña" goat for the first time to enhance its conservation interest. Genomic DNA of seventy five pureblood goats was studied. A PCR-RFLP assay was designed to find a B*my*I target that distinguishing A versus B2 lineages (including recombinant variant M and B1, respectively) of the αS1-casein locus. The allelic frequency of variants related to A lineage (CAG triplet) was 14.0% similar to other southwestern Spanish breeds. It is suggested that individuals or families carrying A lineage should be more studied to detect less allergen null alleles while the opposite allele pools of the B2 lineages should be tested for alleles associated to unsaturated fatty acid content. Therefore, the priorities for conservation plans of animal genetic resources as threatened goat breeds; more investigation is claimed in the aim to study for proved useful alleles of certain genes, as casein variants.

Keywords: αS1-lineage, casein variant, endangered goat breed, genetic resources

1. Introduction

The Retinta (due to show uniform red coat color) Extremeña goat breed is geographically located almost exclusively in Extremadura Autonomy at the Southwestern Spanish territory. This goat is well adapted to "Dehesa" environment and often related to the most depressed socioeconomic uses within their distribution area. Moreover, it has been historically exploited as dual purpose (meat/milk) in a context of familiar subsistence and exclusively under extensive production systems based on browsing and grazing. More recently, the high quality cheese "Queso Ibores" (a Protected Designation of Origin artisan product) used raw milk of this breed and other autochthonous breeds. Despite, its substitution and crossing with more specialized milk yield breeds has occurred since the eighties. This fact led a census reduction of pureblood below 2,000 animals, but in geographically fragmented flocks. From here, the Retinta goat was considered to be on genetic threatening situation reason why the Spanish Agriculture Ministry declared of special protection this breed (R.D. 1682/1997 and R.D. 229/2008). Breeders Association and herd book have been established (Decreto 296/2011) for which the breed was first described with morphological, reproductive and productive data available. However, there is a shallow genetic characterization of the breed in general, much less on casein variants in particular. This is important because this goat is being used to product artisan cheese.

Goat milk synthesized in the mammary gland has six different types of milk proteins (the four caseins -αS1, αS2, β and κ - being 80% of milk protein), among which αS1-casein (αS1-cn: localized to the fourth goat chromosome; Hayes, Petit, Bouniol & Popescu, 1993) has been largely studied due to its impacts and positive correlation with goat milk composition (the amount of total protein, total solids, milk fat concentration and fatty acid composition; see Valenti, Pagano & Avondo, 2012; and references there in) and cheese-making properties (Remeuf, 1993; Pirisi, Colin, Laurent, Scher, & Parmentier, 1994; Clark & Sherbon, 2000; Chilliard et al., 2006). The 18 allelic variants of αS1-cn have been subdivided into four categories according to its quantity in goat milk (Moioli,

D'Andrea, & Pilla, 2007) as follow: (1) high expressing or strong alleles (A, B1, B2, B3, B4, C, H, L and M), (2) intermediate alleles (E and I) and (3) weak alleles (D, F and G) and null alleles (01, 02 and N) with levels of 3.6 g/L, 1.1 g/L and 0.45 g/L (cero for nulls) per allele, respectively (Valenti, Pagano, Pennisi, Lanza & Avondo, 2010). So, strong variants triple the performance levels associated to the αS1-cn and to which should be added the corresponding by the correlation with other milk components already mentioned. Also it has been reported the positive relationship between variants and the physicochemical properties of milk that affects the technological characteristics related to the clotting time of curd, firmness and rennet yield (Grosclaude et al., 1987).

Therefore, genetic studies about casein in threatened goat breed may be reasoned not only to analyze for high yield genes but also to decide about its conservation value based on traits of especial interest for its sustainable use in rural areas. This last statement includes one important of several priorities for conservation of specific animal genetic resources (Boettcher et al., 2010).

The aims of this study was to show for the first time in pureblood "Retinta Extremeña" goat breed valuables casein αS1-lineages to enhance the interest for its conservation.

2. Materials and Methods

Seventy five complete blood samples were collected of the pureblood "Retinta Extremeña" goats from the Selection Centre of Animal Selection and Reproduction (CENSYRA) of the Extremadura Government (4 males and 71 females, around the 45% of the total herd at CENSYRA). Although limited pedigree information was available, only three partial generations was able to assess Mendelian inheritance of the locus. Since all sampled animals belonged to the CENSYRA herd, where this and other endangered breeds are maintained institutionally as pureblood nucleus, the data presented here can be considered to represent this pureblood breed.

Genomic DNA was isolated by non-commercial procedure as in Fernández-García et al. (2012). The identification of natural variants for αS1-cn precursor was obtained in the SWISS-PROT database (CAS1-CAPH1 locus, acc. nº P18626). At 77 (amino acid position) all phylogenetically related casein variants to the most ancestral A variant shows a CAG triplet (glutamine: Q) but other variants wear the triplet GAG (glutamic: E) (Devilacqua et al., 2001). This polymorphism can be identified at 1045-1047 triplet in a segment of the casein coding sequence expands from exon 9 to exon 11 (Leroux, Mazure & Martin,1992) (accession number: GenBank X56462) (Figure 1). Two restriction targets for the endonuclease B*my*I (GDGCH/C) was predicted using software PROPHET (The Prophet Group at BBN Systems and Technologies; http://www-prophet.bbn.com). A primer pair was designed with PRIMER3 software (Koressaar and Remm 2007) that extend partly the intron 9, the exon 10, the intron 10, the exon 11 and partly the intron 11, but flanking both B*my*I target (C-A1 Forward 5´AAGCTATGATGTGTCTGGTT and C-A1Reverse 5´AACATTCTTGCTCATTCCCT). The PCR reaction is performed in 50 µL of a mixture containing template DNA (20-50 ng), primers [10pmol] and other general PCR reagents. Thermal cycling profile was as follow: [97° 5'] / [94ºC - 1'; 59° - 1'; 72° - 1'] * 33 cycles / 72º - 5' final extension; 4º C-∞. 10 µL PCR products were run in agarose gel to assess amplification. After, 10 µL of purified PCR products were digested overnight with 4 U of B*my*I according to the manufacturer **(**Boehringer Mannheim, Germany**).** For detection of genotypes the PCR-RFLP fragments were separated in 2.0 % agarose gel. CERVUS ver 3.0.3 (Marshall, Slate, Kruuk, & Pemberton, 1998) was used to verify HWE for the B*my*I polymorphic site of this locus.

3. Result.

One fragment of 318 bp was always present as expected after PCR assays based on Leroux et al. (1992) sequence (GenBank X56462). Based on knowledge obtained from databases (Figure 1), sequences from the F variant have a 3 bp insertion at intron nine between the selected primers but these should not be resolvable by agarose gels two percent (Figure 1). Digestion with B*my*I revealed the two target sites in the PCR product but only one of them was polymorphic (Figure 2). Accordingly, profiles for three possible genotypes were observed after digestion.

The homozygous goats for the GAG triplet showed profiles consisting of two fragments of 84 and 234 bp (n = 55 genotype GAG/GAG). Therefore, the 234 bp band was present in goats carrying the B2 lineage (see Bevilacqua, Ferrant, Garro, Veltri & Lagonigro, 2002) (Figure 2). In heterozygous goats (n = 19 genotypes GAG/CAG) were seen four bands with 234 bp, 122 bp, 112 bp, but these two last bands almost co-migrant and worst resolved in 2 % agarose gels but signaling the A lineage gene, and 84 bp (Figure 2). The homozygous goats for the CAG triplet have had profiles with 122, 112 and 84 bp bands (n = 1 genotype). Furthermore, the presence of another target for B*my*I was conserved in all samples and therefore it could be used as an internal control of the digestion or enzymatic activity.

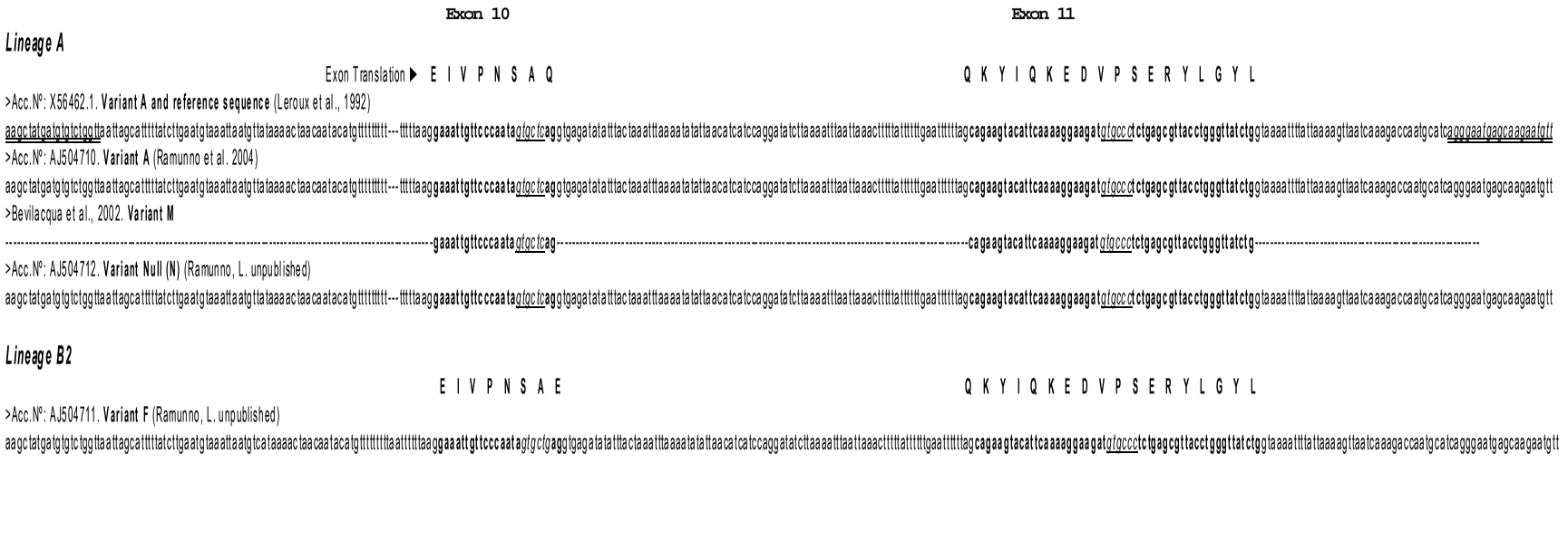

Figure 1. Sequence alignment of different alleles of the αS1-casein.The sequence of the variant A with Accession Number X56462.1 in GenBank (Leroux et al., 1992) was used to design primers and size predictions. Partial sequence of the allele M was transcribed from Bevilacqua et al. (2002). Exon translation: IUPAC code for amino acids of the exon 10 and 11. Double underline for primers sequences, bold type for the nucleotide sequence of each exon, underline-italics for B*my*I target

Co-dominant segregation of B*my*I target was observed in only three parental- descendant available families (data not showed). No discrepant segregation was observed for mendelian transmission of the polymorphic site. The allelic frequency of the CAG triplet was 14.0 % in the total sample, but 37.5 % in the four breeding males. By χ^2 test it was verified HWE equilibrium of the locus using only unrelated animals ($p = 0.801$). So the higher percentage of the GAG triplet in breeding males should be attributed to stochasticity.

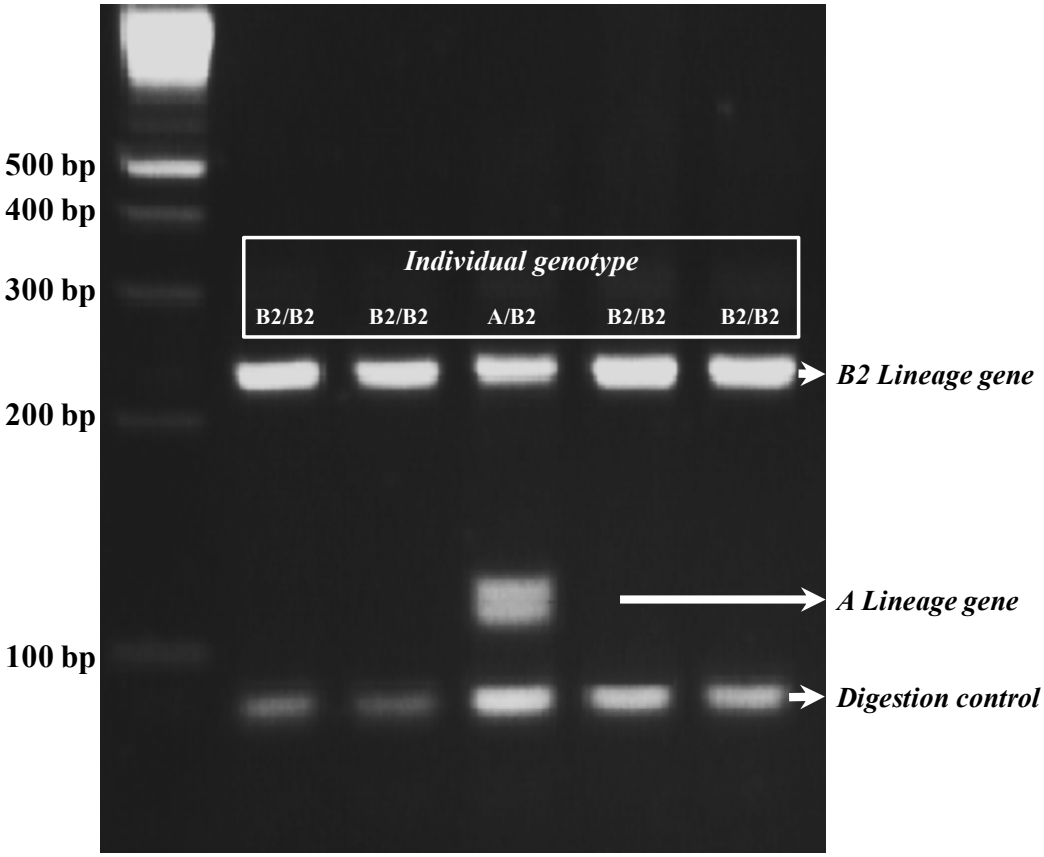

Figure 2. RFLP profile with B*my*I endonuclease from five individuals of the Retinta Extremeña goat. (Right) Restriction fragment corresponding to the gene of each lineage (B2 or A) and the 84 bp fragment digestion Control at the bottom and (left) fragment size of the molecular weight marker

4. Discussion

Structural analysis of goat casein performed both at the protein and the genomic level (Leroux, Martin, Mahè, Leveziel, & Mercier, 1990, Grosclaude & Martin 1997, Sacchi et al., 2005) have showed high variability and complex relation among polymorphic variants for the casein of the goat milk (Mahè & Grosclaude, 1989). This is especially certain for the locus αs1 and κ -casein which showed the highest number of variants (number of variant ≥ 16; Moioli et al., 2007). This high polymorphism feature provides further evidence that the allelic diversity come from multiple pathways, including recombination events between both ancestral lineages (A vs B2) as it has been hypothesized for the αs1- casein locus (Bevilacqua et al., 2002 and references there in)(Figure 3). In this study it was showed that the BmyI-RFLP can be useful to discriminate between lineages A and B2 (including recombinant variant M and B1, respectively) and thus for screening the "allele pools" associated with each one of both in the individual, familiar or population scale (Figure 3). This is important because null alleles (O1, O2 and N) have been exclusively associated to the A lineage (Grosclaude & Martin 1997; Bevilacqua et al., 2002).

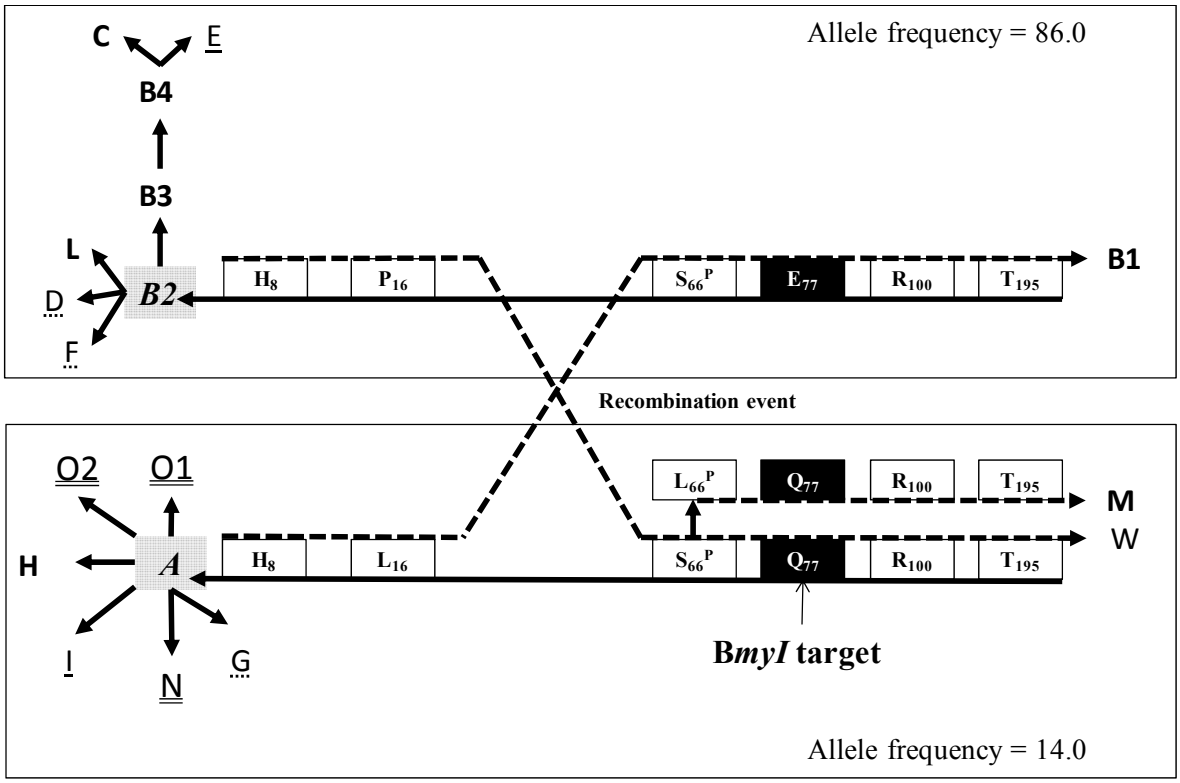

Figure 3. Schematic view of lineages of the αS1-casein locus using a simplified formula for haplotypes. The small boxes mimic exons but in some of them were annotated the polymorphic amino acid residues that occurred between lineages and also it is indicated the amino-acid residue targeted by B*my*I. Both large boxes regroup variants of casein: top for lineage B2 and bottom for lineage A. The numbers at right (down and top) are the frequency of each lineage in "Retinta Extremaña" goat. The continuous lines under the small boxes regroup the amino acid residues belonging to the respective ancestral lineages A or B2 and derived variants but dashed lines are signalling sites of inter-allelic recombination event between both as causes of the B1, W and M variants. Dark arrows indicate the possible evolutionary pathway scenario. Underline of variants: not underlined, simple underline dotted underline and double underlined for those of high yield, average yield, low yield and null (adapted from Bevilacqua et al., 2002)

Several studies revealed that the genotypes with strong or medium alleles of the B2 lineage of the αs1-casein predominated among goat breeds from the Mediterranean range (Sarda, Maltese, Moroccan, Tunisian; Moioli et al., 2007 and references there in) and, within the Iberian Peninsula, in southern breeds as the Murciano-granadina, Malageña or Payoya (Jordana et al., 1996; Caravaca et al., 2009) but allele E being at high frequency (Jordana et al., 1996). Regarding αs1-casein lineages, the genotypic frequencies observed in the Retinta Extremeña breed was similar to other Mediterranean breeds. For example, by pooling alleles belonging to the lineages B2, which carried the GAG triple that coded for the E (glutamic acid) (Bevilacqua et al., 2002), for the murciano-granadina breed (Caravaca et al., 2009) the estimated frequency was around the 88% (the complementary 12% for lineage A). As a whole, the native goat breeds from the Iberian Peninsula showed frequencies of the lineage A within a range between 10% and 20% (Jordana et al., 1996), what includes the results of the Retinta Extremeña reported here. Then, alleles of low frequency but high technological interest might be easily screened for a first glance by the B*my*I-RFLP and thus uncovering hidden αs1-casein lineages on endangered and/or bottlenecked breeds.

Contradictory results about the effect on total protein, fat and casein contents reported for this casein locus (Mahé, Manfredi, Ricordeau, Piacere, & Grosclaude, 1993; Manfredi, Ricordeau, Barbieri, Amigues, & Bibe, 1995; Grosclaude & Martin 1997; Caravaca et al., 2009; Valenti et al., 2010) suggested to be cautious in making decisions about this particular locus regarding the Retinta breed. Both, complicated interactions among genotypes, other *loci* and environmental factors as certain dependence on the particular polymorphism present in populations under consideration should be deemed (Caravaca et al., 2009, Valenti et al., 2010). Recent studies suggested that although some interaction genotype x diet should be expected; comparisons between AA vs FF genotypes affected to nine different fatty acids. But it should stress that monounsaturated fatty acids (especially C16:1 and cis-9 C18:1) were

lower in milk from AA (lineage A) than FF genotypes (lineage B2) (Valenti et al., 2010). Other studies suggested that milk defective in αs1-cn (null alleles) could be less allergenic as reported by Bevilacqua et al. (2001). The goats with these particular genotypes may to be of interest because these have been proposed as substitute of cow milk for people having allergies to cow proteins (Moioli et al., 2007). Regarding to the conservation of goat breeds, all these studies are suggestive because somehow give support to research on the different casein variants within endangered breeds, such as the retina breed. In fact, actions aimed to genotype for proved useful genes should be among the priorities for conservation of specific animal genetic resources (Boettcher et al., 2010), especially for breeds well adapted to difficult territories as the Retinta Extremeña breed. As claimed Caravaca et al. (2009), more investigations in different breeds regarding the effects of casein polymorphism on goat milk yields and the features of their products should be addressed, because it may have an especial value in future comprehensive conservation plans.

Acknowledgements

We thank to CENSYRA (Badajoz, Junta de Extremadura) for providing all samples of the endangered Retinta Extremeña breed and to support this study. We also would like to thank anonymous reviewers for their valuables advices.

References

Bevilacqua, C., Ferrant, P., Garro, G., Veltri C., & Lagonigro, R. (2002). Interallelic recombination is probably responsible for the occurrence of a new αs1-casein variant found in the goat specie. *European Journal* of *Biochemistry, 269*, 1293-1303. http://dx.doi.org/10.1046/j.1432-1033.2002.02777.x

Bevilacqua, C., Martin, P., Candalh, C., Fauquant, J., Piot, M., Roucayrol, A. M., ... Heyman, M. (2001). Goats' milk of defective alpha(s1)-casein genotype decreases intestinal and systemic sensitization to beta-lactoglobulin in guinea pigs. *Journal of Dairy Research, 68*, 217-227. http://dx.doi.org/10.1017/S0022029901004861

Boettcher, P. J., Tixier-Boichard, M., Toro, M., Simianer, H., Eding, H., Gandini, G., … the GLOBALDIV Consortium. (2010). Objectives, criteria and methods for using molecular genetic data in priority setting for conservation of animal genetic resources. *Animal Genetics, 41*(Suppl. 1), 1-14. http://dx.doi.org/10.1111/j.1365-2052.2010.02050.x

Caravaca, F., Carrizosa, J., Urrutia, B., Baena, F., Jordana, J., Amills, M., Badaoui, B., Sanchez, A., Angiolillo, A., & Serradilla, J. M. (2009). Effect of αS1-casein (*CSN1S1*) and κ-casein (*CSN3*) genotypes on milk composition in Murciano-Granadina goats. *Journal of Dairy Science*, 92, 2960-2964. http://dx.doi.org/10.3168/jds.2008-1510

Chilliard, Y., Rouel, J., & Leroux, C. (2006). Goat's alpha-s1 casein genotype influences its milk fatty acid composition and delta-9 desaturation ratios. *Animal Feed Science and Technology, 131*, 474-487. http://dx.doi.org/10.1016/j.anifeedsci.2006.05.025

Clark, S., & Sherbon, J. W. (2000). Alpha S1- casein, milk composition and coagulation properties of goat milk. *Small Ruminant Research, 38*, 123-134. http://dx.doi.org/10.1016/S0921-4488(00)00154-1

Fernández-García, J. L. (2012). The endangered *Dama dama mesopotamica*: genetic variability, allelic loss and hybridization signals. *Contributions to Zoology, 81*(4), 223-233.

Grosclaude, F., & Martin, P. (1997). Casein polymorphisms in the goat. IDF Seminar`Milk Protein Polymorphism II' (pp. 241-253). International Dairy Federation, Palmerston North, New Zealand.

Grosclaude, F., Mahé, M. F., Brigonon, G., Di Stasio, L., & Jeunet, R. (1987). A. Mendelian polymorphism underlying quantitative variation of goat αs1-CN. *Genetics Selection Evolution, 19*, 399-412. http://dx.doi.org/10.1186/1297-9686-19-4-399

Hayes, H., Petit, E., Bouniol, C., & Popescu, P. (1993). Localization of the alpha S2-casein gene (CASAS2) to the homologous cattle, sheep and goat chromosomes 4 by *in situ* hybridization. *Cytogenetics and Cell Genetics, 64*, 281-285. http://dx.doi.org/10.1159/000133593

Jordana, J., Amills, M., Díaz, E., Angulo, C., Serradilla, J. M., & Sanchez, A. (1996). Gene frequencies of caprine αs1-casein polymorphism in Spanish goat breeds. *Small Ruminant Research, 20*, 215-221. http://dx.doi.org/10.1016/0921-4488(95)00813-6

Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics, 23*, 1289-1291. http://dx.doi.org/10.1093/bioinformatics/btm091

Leroux, C., Martin, P., Mahè, M. F., Leveziel, H., & Mercier, J. C. (1990) Restriction fragment length polymorphism identification of the goat as1-casein alleles. A potential tool in selection of individuals

carrying alleles associated with a high level protein synthesis. *Animal Genetics, 21*, 341-351. http://dx.doi.org/10.1111/j.1365-2052.1990.tb01979.x

Leroux, C., Mazure, N., & Martin, P. (1992). Mutations away from splice site recognition sequences might cis-modulate alternative splicing of goat αs1-casein transcripts. Structural organization of the relevant gene. *Journal of Biological Chemistry, 267*, 6147-6157.

Mahè, M. F., & Grosclaude, F. (1989). aS1-CnD, another allele associated with a decreased synthesis rate at the caprine αS1-casein locus. *Genetics Selection Evolution, 21*, 127-129. http://dx.doi.org/10.1186/1297-9686-21-2-127

Mahè, M. F., Manfredi, E., Ricordeau, G., Piacere, A., & Grosclaude, F. (1993). Effets du polymorphisme de la caseine αs1 caprine sur les performances laitieres: Analyse intradescendance de boucs de race Alpine. *Genetics Selection Evolution, 26*, 151-157. http://dx.doi.org/10.1186/1297-9686-26-2-151

Manfredi, E., Ricordeau, G., Barbieri, M. E., Amigues, Y., & Bibe, B., (1995). Genotype caseine αs1 et selection des boucs sur descendance dans les races Alpine et Saanen. *Genetics Selection Evolution, 27*, 451-458. http://dx.doi.org/10.1186/1297-9686-27-5-451

Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology, 7*, 639-655. http://dx.doi.org/10.1046/j.1365-294x.1998.00374.x

Moioli, B., D'Andrea, M., & Pilla, F. (2007). Candidate genes affecting sheep and goat milk quality. *Small Ruminant Research, 68*, 179-192. http://dx.doi.org/10.1016/j.smallrumres.2006.09.008

Pirisi, A., Colin, O., Laurent, F., Scher, J., & Parmentier, M. (1994). Comparison of milk composition, cheesmaking properties and textural characteristics of the cheese from two groups of goats with high or low rate of alfa-s1-casein synthesis. *The International Dairy Journal, 4*, 329-345. http://dx.doi.org/10.1016/0958-6946(94)90030-2

Ramunno, L., Cosenza, G., Rando, A., Illario, R., Gallo, D., Di Berardino, D. & Masina, P. (2004). The goat alphas1-casein gene: gene structure and promoter analysis. *Gene, 334*, 105-111. http://dx.doi.org/10.1016/j.gene.2004.03.006

Remeuf, F. (1993). Influence of genetic polymorphism of caprine αs1-casein on physicochemical and technological properties of goat's milk. *LAIT, 73*, 549-557. http://dx.doi.org/10.1051/lait:19935-652

Sacchi, P., Chessa, S., Budelli, E., Bolla, P., Ceriotti, G., Soglia, D., ... Caroli, A. (2005). Casein haplotype structure in five Italian goat breeds. *Journal of Dairy Science, 88*, 1561-1568. http://dx.doi.org/10.3168/jds.S0022-0302(05)72825-3

Valenti, B., Pagano R. I., & Avondo, M. (2012). Effect of diet at different energy levels on milk casein composition of Girgentana goats differing in CSN1S1 genotype. *Small Ruminant Research, 105*, 135-139. http://dx.doi.org/10.1016/j.smallrumres.2011.11.013

Valenti, B., Pagano, R. I., Pennisi, P., Lanza, M., & Avondo, M. (2010). Polymorphism at αs1-casein locus. Effect of genotype diet × interaction on milk fatty acid composition in Girgentana goat. *Small Ruminant Research, 94*, 210-213. http://dx.doi.org/10.1016/j.smallrumres.2010.07.023

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

Studies on Genetic Diversity of Selected Population of Hybrid Scallop *Chlamys farreri*（♀）× *Patinopecten yessoensis*（♂） by Microsatellites Markers

Biao Wu[1], Aiguo Yang[1], Ningning Cheng [1,2], Xiujun Sun[1], Zhihong Liu[1] & Liqing Zhou[1]

[1] Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, PR China

[2] College of Fisheries and Life Science, Shanghai Ocean University, Shanghai, 201306, PR China

Correspondence: Aiguo Yang, Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao 266071, PR China. E-mail: yangag@ysfri.ac.cn

Received: April 24, 2015 Accepted: May 6, 2015 Online Published: May 15, 2015

doi:10.5539/jmbr.v5n1p39 URL: http://dx.doi.org/10.5539/jmbr.v5n1p39

Abstract

The growth superiority of hybrid scallop *Chlamys farreri* (♀) × *Patinopecten yessoensis* (♂), as the following successive generation selection have been reported. However, the data about the genetic diversity in those population remains unexplored. In this study, the genetic structure analysis of F_1, F_2 and F_3 were conducted by PCR with 10 Simple Sequence Repeats (SSR) primers. It showed that a total of 68 alleles were detected, and the number of alleles per locus ranged from 3 to 11, Polymorphism Information Content (*PIC*) per locus ranged from 0.4729 to 0.8429. And, the average observed heterozygosity (H_o) of the three populations were 0.6100, 0.6975 and 0.7750, while the average expected heterozygosity (H_e) were 0.7607, 0.7751 and 0.7379 respectively. F_{st} values among the three populations were also low (F_{st}<0.05) which suggested low genetic differentiation between each two populations. In all, those data indicated the genetic structure challenge caused by hybridization and selection, supplying a new angle to understand artificial selective breeding.

Keywords: scallop, hybrid breeding, genetic diversity, microsatellite DNA markers

1. Introduction

The scallop *Chlamys farreri*, a native bivalve in China, is one of the most important marine farming species in Northern China. However, massive mortality has caused catastrophic losses to its aquaculture since 1996, which resulted in a sharp production decline. The germplasm quality degeneration have been considered as one major reason for the massive scallop mortality. Therefore, to breed new scallop species with high resistance is an effective method to change the current status quo.

The research on distant hybridization breeding of scallop *C. farreri*(♀) × *Patinopecten yessoensis*(♂) have been carried out (Zhou, Yang, & Liu, 2003; Lv, Yang, Wang, Liu, & Zhou, 2006), and surprisingly, the first hybrid generation owned prominent heterosis performance which has been cultivated in large-scale area. The second and third generation individuals with stronger resistance, faster growth than scallop *C. farreir* have also been selected for farming (Yang, Wang, Liu, & Zhou, 2003). However, during the process of selective breeding, many uncertain factors, such as increasing risk of inbreeding, decreased number of effective groups, may lead to lower genetic diversity, even the genetic effects. Therefore, it is necessary to detect the genetic variation, to understand the changes of genetic structure for developing appropriate scientific measures so that we can smoothly control the progress of selective breeding.

Microsatellite marker (Simple Sequence Repeats, SSR), due to its simple, fast, good stability, higher polymorphism, informative genetic variation and followed Mendelian codominant genetic, has been widely used in various fields as a molecular marker (Li, Park, Endo, & Kijima, 2004; Sakamoto, Danzmann, Okamoto, Ferguson, & Ihssen, 1999; Liu et al., 2004). In terms of hybrid scallop, isoenzyme (He, Yang, Wang, Liu, & Zhou, 2006), RAPD and other labeling techniques were particularly used, however, SSR analysis of different hybrid scallop populations generated from *C. farreri* (♀) × *P. yessoensis* (♂) have not been reported. In this study, SSR was employed to analyze the genetic variation of three selective breeding population, to explore the impact of

selection process on its genetic structure, which could provide a theoretical basis for molecular marker-assisted breeding.

2. Materials and Methods

2.1 Sample Collection

Mature female *C.farreri* and male *P. yessoensis* were collected as parents from Changdao, Shandong Province. The F_1 hybrid were reproduced from mother *C. farreri* and father *P. yessoensis* while F_2 generation from F_1 by self-fertilized, F_3 from F_2, respectively. 30, 40, 40 individuals were selected randomly from F_1, F_2, F_3 population, and stored at -80 °C for DNA extracted, respectively.

2.2 Preparation of Template DNA

The genome DNA was extracted from adductor muscles of hybrid individual using Phenol-chloroform method. In detail, about 100 mg tissue was sampled into a 1.5 ml centrifuge tube, and then 500 μl homogenization buffer (10 mM Tris-Cl, pH 8.0; 100 mM EDTA, pH 8.0), 50 μl 10% SDS and proteinase K with final concentration of 50 μg/ml were also added. After being mixed adequately, the sample were digested at 55 °C for 3 h, and then the proteins were extracted using phenol, phenol: chloroform (1:1), chloroform, isoamyl:alcohol (24:1), successively. And following, the nucleic acid was precipitated with ethanol, and then dissolved in ddH_2O. The concentration and quality of extracted DNA were detected by RNA/DNA quantitative analysis using Nanodrop 2000 and agarose gel electrophoresis, respectively. The concentration of genome DNA was diluted to 50 ng/μl and then stored at -20 °C.

2.3 PCR Amplification

10 pairs of primers (Table 1) were selected from reported SSR primers of *C.farreri* and *P. yessoensis*, to ensure its availability for amplification in hybrid offspring in this study. PCR reactions were carried out in a 10 μl reaction volume on a PCR amplification instrument, including 1 μl 10 × buffer, 0.6 μl Mg^{2+} (25 mM), 1 μl dNTP (each 2 mM), 1 μl forward/reverse primer (10 μM) each, 50 ng template DNA, 0.5 U Taq DNA polymerization enzyme (Promega), and PCR-grade water was replenished to 10 μl. PCR reaction program consisted of 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 72 °C for 30 s, and finally 72 °C for5 min. The PCR products were detected by 6 % denaturing polyacrylamide gel electrophoresis, silver staining for detecting polymorphism. The amplified brands were counted by manual mothed for analyzing by software.

Table 1. Microsatellite primers used in this experiment

Loci	Primer sequence	Annealing temperature	Repeat	GenBank Accession
CFMSM009	F:GTAGTCACATGATGACATAGAG	56	$(AG)_4G(AG)_5\ldots (AG)_5$	DQ104704
	R:CACAACTCCGTCAATCATTCTC			
CFMSM014	F:CATCTGATATGGCAGCTGATAC	60	$(AG)_{10}\ldots(AC)_4$	DQ104705
	R: GAACTAACGAGGAGACAACTG			
CFMSM020	F: CAAAGGCATTTGTAGGAAGGC	62	$(CAC)_{11}$	DQ104709
	R: ACGGCACTTCGTTGATTAAC			
CFMSP011	F: CAAAACCAACTCCTTCACAAC	62	$(ACAAA)_5$	AY682110
	R: GGCGATATTCCACCTGACC			
CFAD213	F: ATTAGTTGTGAAGCAGTCCT	56	$(GA)_{14}$	EF148875
	R:CTTCTCTCAATCATTTCACTATC			
CFFD143	F: CGCCAACCTTGCAGTATCTG	58	$(GA)3CA(GA)_3$	EF148943
	R: TTCTTTCCCTCTTCTGTCCC		$GGAA(GA)_{11}$	
CFBD075	F: TTACTATCCCTACCCCAGAG	60	$(TGTC)_9(TC)_5TG(TC)_{23}$	EF148893
	R:CACTAACCCATTACAAACACAAG		$\ldots (TC)_6 \ldots (CT)_7$	
CFAD053	F:CATTGACACAGTTACAGTTCAC	56	$(CT)_{20}$	EF148859
	R: GCAACAGGATTAGGCACAAG			
S7259	F: CGTCCTTAAATGACCTTA	60	$(AACC)_6$	AY164679
	R: GAAATTCCAGTGTTCGTA			
S9090	F: GAGGAAGAAACATAGTAA	58	$(TTAA)_5$	AY164680
	R: CTACATCAGCTACATCTC			

2.4 Data Analysis

According to the genotype, the value of polymorphic information content (*PIC*), heterozygosity (*Ho*), heterozygosity (*He*), allele number (a), effective number of alleles (a_e), similar coefficient and genetic distances among, and F-statistics of the three groups were calculated using Popgen 32 (Version 1.31). F_{st} value range from 0 to 0.05 was considered as low population genetic differentiation, while 0.05-0.15 as middle level, 0.15-0.25 as high level and above 0.25 as significantly high level.

3. Results

3.1 Genetic Diversity of Loci

10 pairs of primers with high polymorphism were used to perform PCR amplification with 110 individuals from F_1, F_2 and F_3 populations. In all, 68 alleles were obtained, and the numbers of allele for each locus ranged from 3 to 11. *PIC* value was from 0.4729 to 0.8429. The average observed heterozygosity was $F_2>F_1>F_3$, with the values of 0.7025, 0.6893, 0.6900, respectively, while the average expected heterozygosity were 0.8315, 0.7751, 0.7379, respectively. According to *P* values of genotype, the value of the Hardy-Weinberg equilibrium deviated significantly in three generations (as shown in Table 2), such as CFMSM009 in the F_1 and F_2 populations, CFMSP011 in F_3 population and CFAD213 in F_2 and F_3. Some representative amplified brands were exampled in Figure 1.

Table 2. Genetic variability at 10 microsatellite loci in three populations

loci	populations	a	a_e	*PIC*	*Ho*	*He*	*P*-values
CFMSM009	F1	3	2.3407	0.5089	0.3333	0.5825	0.0068**
	F2	3	2.0901	0.4613	0.2750	0.5282	0.0146*
	F3	3	2.0566	0.4380	0.1500	0.5203	0.0000**
CFMSM014	F1	7	4.6512	0.7613	0.6667	0.7983	0.2517
	F2	7	5.9590	0.8110	0.7000	0.8427	0.0942*
	F3	6	4.1995	0.7244	0.7000	0.7715	0.3040
CFMSM020	F1	11	7.2581	0.8468	0.7333	0.8768	0.2247
	F2	10	6.8085	0.8365	0.9000	0.8963	0.4101
	F3	10	6.4000	0.8258	0.9750	0.8544	0.0823*
CFMSP011	F1	7	3.7657	0.6956	1.0000	0.7469	0.0110*
	F2	7	4.3096	0.7327	0.9750	0.7775	0.0162*
	F3	6	3.2196	0.6387	0.9737	0.6986	0.0001**
CFAD213	F1	6	3.1304	0.6363	0.4667	0.6921	0.0315*
	F2	6	3.7915	0.6964	0.3750	0.7456	0.0001**
	F3	6	2.6801	0.5663	0.3750	0.6348	0.0005**
CFFD143	F1	8	4.6036	0.7509	0.8667	0.7960	0.2007
	F2	7	4.1290	0.7227	0.8500	0.7674	0.4974
	F3	6	3.8508	0.7009	0.8000	0.7497	0.7390
CFBD075	F1	6	4.2857	0.7328	0.6333	0.7797	0.0106*
	F2	6	3.6036	0.6885	0.6750	0.7316	0.2184
	F3	6	3.1809	0.6466	0.5500	0.6943	0.0290*
CFAD053	F1	8	6.7924	0.8213	0.7667	0.7469	1.0000
	F2	7	6.7368	0.8331	0.7000	0.8623	0.0158*
	F3	8	5.9530	0.8118	0.7692	0.8423	0.4489
S7259	F1	6	4.9451	0.7670	0.4333	0.8113	0.3268
	F2	6	4.9231	0.7664	0.5750	0.8070	0.0392*
	F3	6	4.3716	0.7390	0.6000	0.7810	0.0837
S9090	F1	6	4.2254	0.7279	1.0000	0.7763	0.0280*
	F2	6	5.3872	0.7873	1.0000	0.8247	0.1639
	F3	6	5.5817	0.7957	1.0000	0.8315	0.1643
Mean	F1	6.8000	4.5998	0.7241	0.6900	0.8315	
	F2	6.5000	4.7736	0.7336	0.7025	0.7751	
	F3	6.2000	4.1494	0.6930	0.6893	0.7379	

Note: * indicates significant deviation (P <0.05); ** indicates highly significant deviation (P <0.01).

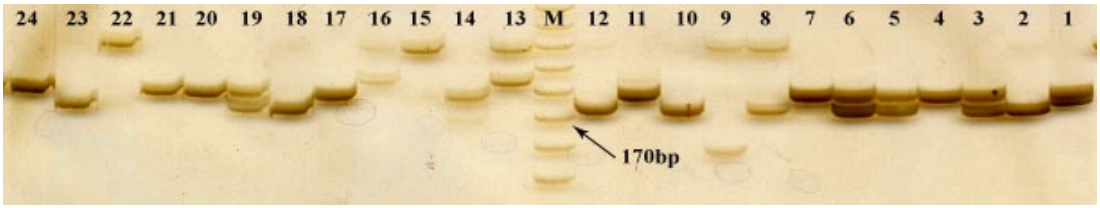

Figure 1. Demonstration of microsatellite locus amplified by CFMSM020 primer pairs in three populations
M: Marker; 1-8: F_1 population; 9-16: F_2 population; 17-24: F_3 population

3.2 Genetic Similarity Index, Genetic Distance and Cluster Analysis

The genetic distance and genetic similarity in three population was shown in Table 3. It showed that the genetic distance between F_1 and F_3 generation was the largest, while the value between F_2 and F_3 populations was the smallest. According to the genetic distance among groups, UPGMA were used to analyze relationship among three generations populations (Figure 2). It indicated that the F_2 and F_3 clustered firstly and then cluster with F_1.

Table 3. Genetic identity and genetic distance in three populations

Population	F_1	F_2	F_3
F_1	—	0.9465	0.9176
F_2	0.055	—	0.9716
F_3	0.086	0.0288	—

Notes: data below diagonal are genetic distance; data above diagonal are genetic identity.

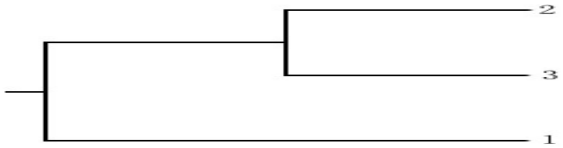

Figure 2. UPGMA dendrogram among three populations
1: F_1 population; 2: F_2 population; 3: F_3 population

3.3 Population Variation

The F_{st} values between F_1 and F_2, F_1 and F_3, F_2 and F_3 were 0.0079, 0.0196, 0.0028, respectively, which indicated that the genetic differentiation among the three generations was weak. The overall genetic differentiation coefficient value was 0.0169, which revealed that only 1.69 % genetic variation was from groups, while 98.31% variation was from individuals.

Table 4. *F*-statistica for three populations of hybrid scallop at ten microsatellite

loci	F_{is}			F_{st}
	F1	F2	F3	
CFMSM009	0.4180	0.4727	0.7080	0.0082
CFMSM014	0.1507	0.1588	0.0812	0.0148
CFMSM020	0.1495	-0.0549	-0.1556	0.0075
CFMSP011	-0.3616	-0.2698	-0.4124	0.0135
CFAD213	0.3143	0.4907	0.4018	0.0376
CFFD143	-0.1072	-0.1216	-0.0806	0.0235
CFBD075	-0.1739	0.0657	0.1978	0.0090
CFAD053	-0.0439	0.1780	0.0755	0.0355
S7259	0.4568	0.2784	0.2220	0.0042
S9090	-0.3100	-0.2279	-0.2180	0.0132
mean	0.0775	0.0822	0.0539	0.0169

Table 5. F_{st} values of pairwise comparison among different population at 10 microsatellite

Groups	F_1	F_2	F_3
F_1	—	0.0079	0.0196
F_2		—	0.0028
F_3			—

4. Discussion

In this study, the data revealed that the genetic diversity among three populations was not significantly different, although the proportion of polymorphic loci and genetic diversity decreased from F_1 to F_3. This situation was similar to some other reports (Hedgecock, Chow, & Waples, 1992; Mgaya, Gosling, Mercer, & Donlon, 1995; V. Sbordoni, De Matthaeis, M. C. Sbordoni, La Rosa, & Mattoccia, 1986). For example, the genetic diversity in several generations of Chinese shrimp were reported by Zhang et al. (2005) using SSR technology, which showed that the average observed heterozygosity dropped down from the first generation to the sixth. And it was same as the studied in American oysters (Yu and Guo, 2004) and Japanese flounder (Liu et al. 2005). Selective breeding is a complex process, the external environment and artificial selection pressure may cause fluctuations of population genetic variation. And on the other hand, during the artificial breeding process, due to the small effective population the inbreeding rate might increase which caused inbreeding depression and bottleneck effect. And also, high-intensity artificial directional selection might lead to the genetic deterioration and introgression, that these two factors could cause the loss of some particularly alleles, especially some rare gene allele in the breeding population. Therefore, genetic variation of breeding populations should be detected timely in the process of selective breeding.

Coefficient of genetic differentiation is an important parameter reflecting the degree of genetic differentiation among populations. Under controlled conditions, artificial selection, mutagenesis, hybridization could damage balance of genetic, which caused changes of genes and genotype, so the genetic characteristics within a population will also change. In this study, from the first generation to the third generation of breeding populations, the genetic structure within populations changed. Although the genetic structure had a certain differentiation, the degree of differentiation was not significant. The genetic variation analysis showed that genetic differentiation among generations of artificial breeding populations was smaller and differentiated was mainly from individuals. As studies in Chinese shrimp reported by Li et al. (2006), the genetic differentiation coefficient between adjacent groups showed a decreasing trend, and genetic similarity degrees of individuals within a population showed an upward trend with the increase in generation, which indicated the breeding populations tend stability after years of breeding.

Heterozygosity is an important parameter to measure the genetic diversity of populations. In this study, the mean observed heterozygosity were from 0.6893 to 0.7025 among the three generations. High significant deviation phenomena of Hardy-Weinberg equilibrium were observed, and the F_{is} values indicated that seven loci in three populations showed a certain degree loss of heterozygosity, which indicated dumb allele might exist, which was similar to the report of Chinese shrimp (Zhang et al., 2005)). This study enriches the data of hybrid breeding of scallop, opening a new angle to understand genetic diversity change of different generation during artificial selective breeding.

References

He, B., Yang, A. G., Wang, Q. Y., Liu, Z. H., & Zhou, L. Q. (2006) . A comparative study on isozymes in the hybrids of *Chlamys farreri* (♀) × *Patinopecten yessoensis* (♂) and their parent stocks. *Marine Fisheries Research, 27*(5), 23-27.

Hedgecock, D., Chow, V., & Waples, R. S. (1992). Effective population numbers of shellfish broodstocks estimated from temporal variance in allelic frequencies. *Aquaculture, 108*(3), 215-232. http://dx.doi.org/10.1016/0044-8486(92)90108-W

Li, Q., Park, C., Endo, T., & Kijima, A. (2004). Loss of genetic variation at microsatellite loci in hatchery strains of the Pacific abalone (*Haliotis discus hannai*). *Aquaculture*, *235*(1), 207-222. http://dx.doi.org/10.1016/j.aquaculture.2003.12.018

Li, Z., Li, J., Wang, Q., He, Y., & Liu, P. (2006). The effects of selective breeding on the genetic structure of shrimp *Fenneropenaeus chinensis* populations. *Aquaculture, 258*(1), 278-282.

Liu, P., Meng, X. H., He, Y. Y., Kong, J., Li, J., & Wang, Q. Y. (2004). Genetic diversity in three wild populations of shrimp *Fenneropenaeus chinensis* in Yellow and Bohai Seas as revealed by microsatellite DNA. *Oceanologia et limnologia sinica, 35*(3), 257-262.

Liu, Y., Chen, S., & Li, B. (2005). Assessing the genetic structure of three Japanese flounder (*Paralichthys olivaceus*) stocks by microsatellite markers. *Aquaculture, 243*(1), 103-111. http://dx.doi.org/10.1016/j.aquaculture.2004.10.024

Lv, Z. M., Yang, A. G., Wang, Q. Y., Liu, Z. H., & Zhou, L. Q. (2006). Preliminary cytological identification and immunological traits of hybrid scallop from *Chlamys farreri* (♀) × *Patinopecten yessoensis* (♂) [J]. *Journal of Fishery Science of China, 13*(4), 597-602.

Mgaya, Y. D., Gosling, E. M., Mercer, J. P., & Donlon, J. (1995). Genetic variation at three polymorphic loci in wild and hatchery stocks of the abalone, Haliotis tuberculata Linnaeus. *Aquaculture, 136*(1), 71-80. http://dx.doi.org/10.1016/0044-8486(95)01037-8

Sakamoto, T., Danzmann, R. G., Okamoto, N., Ferguson, M. M., & Ihssen, P. E. (1999). Linkage analysis of quantitative trait loci associated with spawning time in rainbow trout (Oncorhynchus mykiss). *Aquaculture*, *173*(1), 33-43. http://dx.doi.org/10.1016/S0044-8486(98)00463-3

Sbordoni, V., De Matthaeis, E., Sbordoni, M. C., La Rosa, G., & Mattoccia, M. (1986). Bottleneck effects and the depression of genetic variability in hatchery stocks of Penaeus japonicus (Crustacea, Decapoda). *Aquaculture, 57*(1), 239-251. http://dx.doi.org/10.1016/0044-8486(86)90202-4

Yang, A. G., Wang, Q. Y., Liu, Z. H., & Zhou, L. Q. (2003). The hybrid between the scallops *Chlamys farreri* and *Patinopecten yessoensis* and the inheritance characteristics of its first filial generation. *Marine Fisheries Research*, *25*(5), 1-5.

Yu, Z., & Guo, X. (2004). Genetic analysis of selected strains of eastern oyster (*Crassostrea virginica Gmelin*) using AFLP and microsatellite markers. *Marine biotechnology*, *6*(6), 575-586. http://dx.doi.org/10.1007/s10126-004-3600-5

Zhang, T. S., Wang, Q. Y., Liu, P., Li, J., & Kong, J. (2005). Genetic diversity analysis on selected populations of shrimp *Fenneropenaeus Chinensis* by microsatellites. *Oceanologia Et Limnologia Sinica, 29*(1), 6-12.

Zhou, L. Q., Yang, A. G., & Liu, Z. H. (2003). Cytological Observations on Cross Fertilization between *Chlamys farreri (♀) and Patinopecten yessoensis (♂)* Scallops. *Chinese Journal of Zoology Peking, 38*(4), 20-23.

Zhou, L. Q., Yang, A. G., Liu, Z. H., Du, F. Y., & Wang, Q. Y. (2003). Electron microscope observation on *Patinopecten yessoensis* sperm penetration into Chlamys farreri egg. *Journal of Fishery Sciences of China, 3*, 002.

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

Real Time PCR: the Use of Reference Genes and Essential Rules Required to Obtain Normalisation Data Reliable to Quantitative Gene Expression

Antônio J. Rocha[1], José E. Monteiro-Júnior[2], José E.C. Freire[1], Antônio J.S. Sousa[1] & Cristiane S.R. Fonteles[3]

[1] Departamento de Bioquímica e Biologia Molecular, Avenida Humberto Monte, s/n - Pici - CEP 60440-900, Fortaleza - CE, Brasil

[2] Departamento de Biologia, Avenida Humberto Monte, s/n - Pici - CEP 60440-900, Fortaleza - CE, Brasil

[3] Departmento de Clínica Odontológica, Universidade Federal do Ceará, Rua Monsenhor Furtado, s/n - Rodolfo Teófilo - CEP 60430-350, Fortaleza - CE, Brasil

Correspondence: Antônio J. Rocha, Departamento de Bioquímica e Biologia Molecular, Avenida Humberto Monte, s/n - Pici - CEP 60440-900, Fortaleza - CE, Brasil. E-mail: antonionubis@gmail.com

Received: July 1, 2015 Accepted: July 15, 2015 Online Published: July 20, 2015

doi:10.5539/jmbr.v5n1p45 URL: http://dx.doi.org/10.5539/jmbr.v5n1p45

Abstract

Quantitative Real-time Polymerase Chain Reaction (qPCR) is an important tool for molecular biology and biotechnology research, widely used to determine the expression levels of mRNA. Two main methods to performing qPCR are largely used: The absolute quantification, in which the mRNA levels are determined by using a standard curve and the relative method, which is based on the use of reference genes. Reference genes are widely expressed in cells of animal and plant tissues and their expression pattern are theoretically unchanged within several situations, which makes them an excellent choice to normalize mRNA quantification data in relative qPCR studies. However, several reports are increasingly showing that the use of only one reference gene in relative qPCR studies should be avoided, because in the real world their expression levels can significantly change from tissue to tissue. Several softwares, such as geNorm, BestKeeper and NormFinder, have been developed to perform data normalisation, and these programs may assist in choosing the most stable reference genes. The aim of this review was to describe the current normalisation strategies used in qPCR assay, as well as to establish essential rules to perform reliable mRNA quantification. Finally, this review show some innovations in the advances on qPCR.

Keywords: primer design, DNA binding dyes, probes, normalisation

1. Introduction

The polymerase chain reaction (PCR) technique was first introduced by Kary Mullis (Saiki et al., 1985). PCR is historically used as a sensitive method for the detection and amplification of specific sequences of nucleic acids in a sample. Advances in the specificity and sensibility of PCR reactions gave birth to a more sensitive PCR technique, namely quantitative PCR (qPCR-quantitative real-time polymerase reaction), which utilizes mainly cDNA as template, a complementary DNA from RNA molecules through of the reverse transcriptase reaction. In these reactions, fluorescent reporters used include double-stranded DNA (dsDNA) binding dyes or probes that are incorporated into the product during amplification. The increase in fluorescent signal is directly proportional to the number of PCR product molecules generated in the reaction.

qPCR is amongst the best available methods to determining changes in gene expression, due their ability to rapidly and accurately quantify target genes, even in the presence of very low expression levels (Holland, 2002). Prior to the analysis of gene expression, the selection of an appropriate normalisation strategy is essential to control for non-specific variations between samples of cDNA. The most commonly used method to normalising qPCR data is relied on the use of one or more endogenous reference genes (Hamalainen et al., 2001; Rebouças et al., 2013).

Reference genes have uniform and stable expression in a wide variety of tissues and cell types, at different developmental stages, and comprise all genes that express protein products involved in basic cellular processes

(Reid et al., 2006), showing none or only minimal changes in the expression levels between individual samples and experimental conditions (Rebouças et al., 2013). These genes are largely used as internal controls for normalisation in gene expression studies in different tissues and/or condition as in plants and animals (Wong et al., 2005; Kumar et al., 2013; Sara et al., 2013; Rocha et al., 2013; Nakayama et al., 2004). Several reference genes, including those coding for biological products such as tubulins, actin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), phosphatases, albumin, cyclophilin, micro-globulin, ribosomal units (18S rRNA) or ubiquitin (UBQ) have been described in the literature (Foss et al., 2003; Rocha et al., 2013). The correct choice of reference genes is crucial to properly analyze the results of qPCR (Suzuki et al., 2000) and to measure and reduce the errors from variations among the samples (Barsalobres-Cavallari et al., 2009).

Several research groups have developed software tools to identify the most stable expressed genes across a set of samples in order to perform data normalisation. These tools include geNorm, NormFinder and BestKeeper (Vandesompele et al., 2002; Pfaffl et al., 2004; Andersen & Orntoft, 2004), which is freely available on the web and allows researchers to find the best reference gene for their experiments. A great number of studies describing the identification of multiple reference genes for normalisation of qPCR data using these algorithms have been performed on the animal and human health fields (Hong et al., 2008; De Boever et al., 2008), but similar reports are scarce in plant research (Jain et al., 2006; Ransbotyn et al., 2006; Exposito-Rodriguez et al., 2008).

The aim of this review was to evaluate the importance of the application of reference genes in normalisation strategies of qPCR assays, in different tissues or experimental conditions, as well as to describe essential rules necessary to conduct successful qPCR experiments. Besides, we pointed out several precautions required for a good qPCR. Finally, this review shows some innovations in the advances on qPCR in the last years.

2. DNA Binding Dyes Versus Hydrolysis Probes in qPCR

PCR is one of the most versatile technologies in molecular biology. The PCR reaction consists of 3 different stages which involve, (a) the DNA denaturation; (b) the primer annealing and (c) the extension phase (Mullis et al., 1987). In traditional (endpoint) PCR, the detection and quantification of amplified target sequences are performed at the end of the reaction, and it involves additional work, such as gel electrophoresis and image analysis. Nevertheless, in qPCR, the amount of PCR product is measured along each reaction cycle. The ability to monitor the reaction during its exponential phase enables users to determine the initial amount of target gene with great precision (Wong et al., 2005).

In qPCR, the amount of DNA is measured by the use of fluorescent markers that are incorporated into the PCR product. The increase in fluorescent signal is directly proportional to the number of PCR product molecules (amplicons) generated in the exponential phase of the reaction. Fluorescent reporters used include double-stranded DNA (dsDNA) binding dyes or probes that are incorporated into the product during amplification (Bustin et al., 2002). SYBR Green is an example of a fluorescent dye which binds to the double-stranded DNA and emits light upon excitation. Once the reaction proceeds and the PCR product is accumulated, the fluorescence levels increase proportionally to the amount of DNA present in the original sample (Livak et al., 1995; Pabla et al., 2008; Bustin et al., 2002). This dye is used to monitor the amplification of any DNA sequences and dispenses the use of a probe, thus reducing the cost of amplification and providing a great advantage in its application. On the other hand, since the dye binds not only to the target DNA, but to all dsDNA formed during qPCR, the use of SYBR Green, while simple lacks specificity (Figure 1a). The specificity of the reactions, however, can be easily accessed by the use of melting curve analysis (Dheda et al., 2004).

In addition to DNA binding dyes, there are probes, such as TaqMan®, which are designed to binds to specific DNA sequences. TaqMan® probes primarily consist in a oligonucleotide sequence complementary to some regions of the target DNA. The probe is complexed with a quencher and a reporter fluorophore dye at its 3' and 5' ends, respectively (Livak et al., 1995). During the amplification step the probe is associated to its complementary target DNA and then is cleaved by *Taq* DNA polymerase 5'-3' exonuclease activity (Figure 1b). This cleavage releases the reporter dye and generates a fluorescent signal that increases with each cycle (Bustin et al., 2002). TaqMan® provides higher specificity than DNA intercalating dyes, such as SYBR Green. In addition, these probes can also be labeled with distinct and distinguishable reporter dyes, which allows the amplification of two different sequences in the same reaction tube, eliminating the post-PCR processing, and reducing hand labor. The main drawback of this system is the requirement to synthesize specific probes to each target sequences, increasing the cost of the assay (La Cruz et al., 2013).

Another type of probe which is largely used in qPCR assay is molecular beacon. When free in solution molecular beacon probes assume a hairpin structure consisting of a quencher and a reporter dye (Tyagi et al., 1996). The reporter fluorescent dye and the quencher remain extremely close and therefore no fluorescence is detected when

this structure is formed (Figure 1c). However, during the annealing step, Molecular Beacon hybridizes to the target sequence generating conformational changes leading to the separation of reporter and quencher dyes, which results in the emission of fluorescence (Tyagi et al., 1996; VanGuilder et al., 2008). The greater specificity for mismatch discrimination is due to structural constraints. However, the main disadvantage associated with Molecular Beacons is the accurate design of the hybridization probe. Optimal design of the Molecular Beacon stem annealing strength is crucial (Wong et al., 2005).

Scorpions consist of a single-stranded oligonucleotide probe of approximately 20 to 25 nt carrying a reporter fluorophore at its 5' end and a quencher at its 3' end. Their tridimensional conformation resembles a stem and loop structure, in which a PCR primer is attached (figure 1d). The stem-and-loop structure acts as a blocker to prevent DNA polymerase activity during the interaction of the probe with the target DNA (Bustin et al., 2002; Ng et al., 2005). The close proximity of the reporter to the quencher leads to a continuous suppression of the fluorescence emitted by the reporter. At the beginning of the PCR, *Taq*DNA polymerase extends the PCR primer and synthesizes the complementary strand of the target sequence (Whitcombe et al., 1999). During the next cycle, the stem-and-loop structure unfolds and the loop region of the probe hybridizes intra-molecularly to the newly synthesized target sequence. The reporter is excited by light from the qPCR instrument (Bustin et al., 2002). Once the reporter dye is no longer in close proximity to the quencher dye, fluorescence emission may take place. The significant increase of the fluorescent signal is detected by the qPCR instrument and it is directly proportional to the amount of target DNA (Holland, 2002; Wong et al., 2005; Kumar et al., 2013). Scorpions have the advantage to providing a stronger signal and lower level of background when in compared to other probes, such as molecular beacons (Bustin et al., 2002).

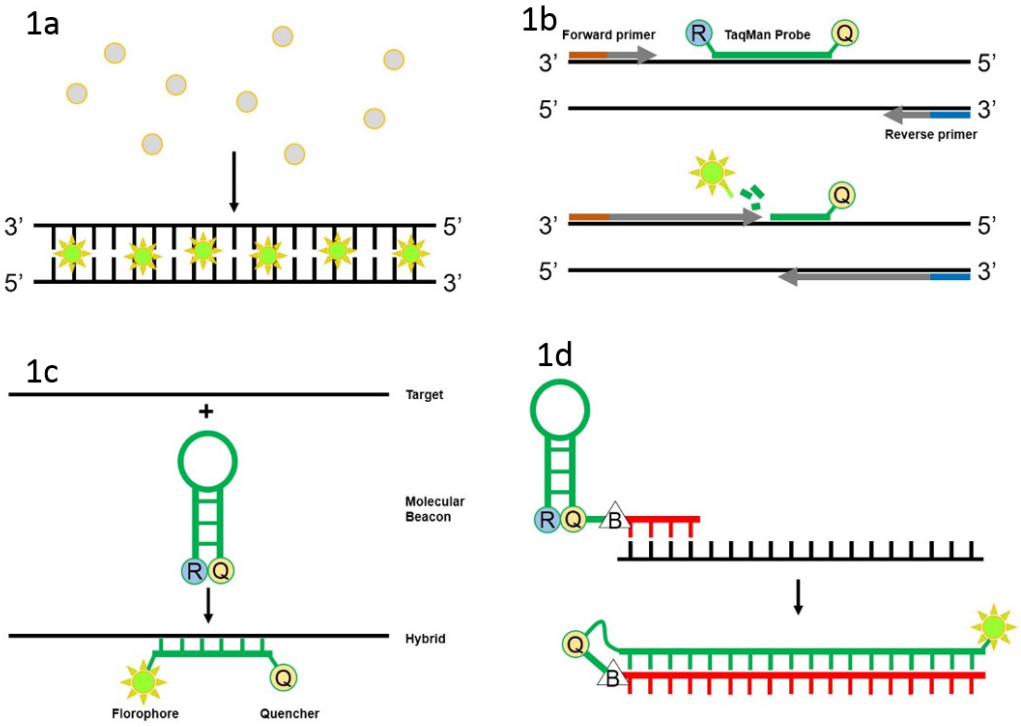

Figure 1. Probes and Dyes used in Real time PCR assay; a) SYBR Green; b) TaqMan; c) Molecular Beacon; d) Scorpions

3. The Use of Reference Genes to Normalize qPCR Data

Reference genes in qPCR are critical for normalisation of expression levels, thus, avoiding misinterpretation of results obtained by real time PCR data. In recent years, it has become clear that no single gene is constitutively expressed in all cell types and under all experimental conditions. This implies that the expression stability of a putative control gene (reference gene) must be verified before each qPCR assay and that the use of only one reference gene is generally not enough to normalize the expression data (Livak et al., 2001; Lee et al., 2010).

The choice of several reference genes to normalize and validate the final results may significantly influence the accuracy of gene expression. Consequently, the use of inappropriate reference genes for normalisation of expression data may lead to erroneous results and data misinterpretation (Suzuki & Higgins, 2000), because normalisation is a pivotal step that provides the Cq values-based differences between the reference and target genes, avoiding misinterpretation of the results and providing reliable Cqs, thus rendering a more accurate and reliable gene expression (Vandesompele et al., 2002). The Ct or threshold cycle value is the cycle number at which the fluorescence generated within a reaction crosses the fluorescence threshold, a fluorescent signal significantly above the background fluorescence. Therefore, the selection of appropriate reference genes is a critical step before evaluating gene expression in new species and/or tissues (Condori et al., 2001; Cordoba et al., 2001). The best candidate genes are those selected by programs used to establish reference genes, such as geNorm, BestKeeper and NormFinder. Therefore, the normalisation using appropriated reference genes are pivotal to acquire suitable data and avoid and misinterpretation of the experiments.

4. Algorithms Used to Normalize qPCR Data

In the last decade, relevant tools to select genes for normalisation have become available. Several research groups have developed softwares to identify the most stably expressed genes across a set of samples. Among these tools we will focus on the most cited articles as geNorm, NormFinder and Bestkeeper (Vandesompele et al., 2002; Andersen & Orntoft, 2004; Pfaffl et al., 2004), which are freely available on the web and allow researchers to find the best reference gene for their experiments. These programs allow the calculation of a normalisation factor over multiple reference genes, which improve the robustness of the normalisation even further (Dekkers et al., 2012). Different manners to access the stability of putative reference genes are available using the upon mentioned software. Hence, BestKeeper employs quantification cycle (C_q) values directly for stability calculations, whereas geNorm and NormFinder have these values transformed to relative quantities using normalisation factor (NF) (Mallona et al., 2004).

3.1 Genorm Analysis

The geNorm program has been recently reported to be one of the best statistical methods to identify stably expressed genes for qPCR analysis. The geNorm calculates a gene-stability measure *M* as the Average pairwise variation *V* of a particular gene reported to all other control genes. Genes with the lowest *M* values have the most stable expression. Stepwise exclusion of the gene with the highest *M* value allows ranking of the tested genes according to the stability (Condori et al., 2001; Cordoba et al., 2001; Vandesompele et al., 2002; Zhong et al., 2009). The analysis relies on the principle that the expression ratio of two proper control genes should be identical in all samples, regardless of the experimental conditions or cell type, and the M values below cutoff (< 1.5) are regarded the most stable genes among all candidate reference genes (Vandesompele et al., 2002).

The geNorm program estimates also the number of genes required to be used as appropriate controls for normalisation by evaluation of variation in pairs (V values), checking the variation of the expression of two by two possible genetic combinations . The optimal number of reference genes that should be used for accurate normalisation also depends on the specific experimental condition, which is determined by calculating V-values as a pairwise variation (Vn/Vn+1) between two consecutively ranked normalisation factors (NF) after the stepwise addition of the subsequent more stable reference gene (NFn and NFn+1) (Vandesompele et al., 2002). Actually, the geNorm is part of qBASEPlus (Biogazelle) program as tool important to provide the reference genes more stables (M value) and the number of genes suitable to normalisation (V value). Furthermore, the qBASEPlus (Biogazelle) also provides the relative expression on qPCR experiments based on the normalisation factor (NF). The use of the qBASEPlus (Biogazelle) is needed at least 8 reference genes and at least 2 samples (control and conditions) for to analyze the qPCR data.

3.2 Normfinder Analysis

The NormFinder is an algorithm used to identify the optimal normalisation gene among a set of candidates. It ranks the set of candidate normalisation genes according to their expression stability in a determined sample set and given experimental design. This algorithm is rooted in a mathematical model of gene expression and uses a solid statistical framework to estimate not only the overall expression variation of the candidate normalisation genes, but also the variation between sample subgroups of the sample set e.g. normal and cancer samples (Andersen & Orntoft, 2004). Notably, "NormFinder" provides a stability value for each gene, which is a direct measure for the estimated expression variation, enabling the user to evaluate the systematic error introduced when using the gene for normalisation (Dekkers et al., 2012; Selim et al., 2012).

3.3 Bestkeeper Analysis

The BestKeeper calculates standard deviation (SD) and the coefficient of variation (CV) based on Cq values of all reference candidate genes. Genes with SD less than 1 are considered stable. Subsequently, the program calculates a pairwise correlation coefficient between each gene and the BestKeeper index–geometric mean between Ct values of stable genes grouped together. Genes with the highest coefficient of correlation with the BestKeeper Index indicates the highest stability (Pfaffl et al., 2004). The BestKeeper use raw Ct data and determines the most stably expressed genes based on a correlation coefficient (r) of the BestKeeper Index (BI) and standard deviation, whereas BI is the geometric mean of Ct values of best reference genes. Hence, this program relies on the “r” and “SD” values, and the higher the "r" value, the most stable is the gene; otherwise, the lower the standard deviation value, the most stable is the gene (Pfaffl et al., 2004; Demidenko et al., 2011; Niu et al., 2011; Petit et al., 2012).

These statistical algorithms have been developed for the evaluation of best suited reference gene(s) for normalisation of qPCR data in a set of biological samples. Recognizing the importance of reference genes in normalisation of qPCR data, various reference genes have been evaluated for their stable expression under specific conditions in various organisms. Many studies have been conducted in the animal and human health (De Boever et al., 2008; Hong et al., 2008) fields that describe the identification of multiple reference genes for normalisation of qPCR data, but similar reports are scarce in plant research (Ransbotyn et al., 2006; Exposito-Rodriguez et al., 2008).

The three algorithms are important for reference gene stability and normalisation data during qPCR experiments; however, geNorm is the best tool since in addition to providing the best reference genes in geNorm M, this software supplies the V value, which delivers the number of genes needed for use in normalisation data in a qPCR experiment. The algorithms NormFinder and BestKeeper will only identify the most stable genes. Generally, all three algorithms are used to render more reliable results for normalisation.

5. Essential Rules Required to Perform a Reliable qPCR

The efficiency and specificity of quantitative PCR depends on several parameters related to quantification of mRNA, which must be controlled to avoid errors of interpretation: purification of RNA, efficiency of primer specifics, normalisation of reference genes, tissue inhibitory factors, enzyme loading error (Rocha, Miranda, & Cunha, 2014), pipetting errors, among others (Thellin et al., 1999; Livak et al., 2001; Suzuki et al., 2000; Vandesompele et al., 2002) as described below.

5.1 Primers and Probes Design Considerations

The primers and probes design are essentials for amplification efficiency, specificity and fluorescence, respectively. The primers are specially needed in junction exon-exon to avoid an amplification of DNA genomic, ensuring the amplification of only a target gene specific cDNA sequence. In addition, it may be necessary to digest input DNA with an RNase free DNA in the following circumstances: (1) to avoid DNA amplification during qPCR; (2) to use primers that either flank an intron that is not present in the mRNA sequence or that span an exon-exon junction; (3) when the gene of interest has no introns; (4) if the intron positions are unknown; (5) when there are no suitable primers that span or flank introns (Udvardi et al., 2004). There are several programs used to design automatic primers, such as Perl Primer (Marshall, 2004) that will require previous annotation of genes, establishing the introns and exons of each sequence to input the program. Other programs are available in the web as primer BLAST, a tool available at http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_ LOC=BlastHome in the GenBank of NCBI, as well as Primer3 Plus available at http://www.bioinformatics.nl/ cgi-bin/primer3plus/primer3plus.cgi/. Moreover, an absence of primer-dimer and non-specific amplification is especially important to suitable data of qPCR. Therefore, the presence of homo-dimers, hetero-dimers, as well as self-dimers must be avoided, and the formation of harpin of a forward or reverse primer (Condori et al., 2001; Rocha et al., 2013).

The probes, such as TaqMan®, Molecular beacon and scorpions are primers marked with fluorophores to emit fluorescence. These probes are designed in different forms, but are used with the common purpose of emitting fluorescence to assess the increase on gene expression due to the number of probes that bind a double-stranded DNA (Bustin et al., 2002; Pabla et al., 2008; VanGuilder et al., 2008; Hwang et al., 2013). There are programs, such as primer BLAST, available in the GenBank of NCBI, as well as Primer3 Plus both available in the web. These are the same programs used to design primers and probes (Condori et al., 2001).

5.2 RNA Quality

The quality of RNA also is very important to provide accurate qPCR data. The quality of RNA depends on extraction and purification of RNA, for example, during the extraction of RNA contaminants such as proteins,

carbohydrate, as well as phenolic compounds that will affect the PCR reaction by inhibiting the action of polymerases as reverse transcriptase and DNA polymerases, during qPCR must be avoided. Therefore, RNA of good quality is needed for further experiments. According to Sambrook et al. (1989), the best relations absorbance by spectrophotometer are as follows: RNA relation to $A_{260/280}$: 1.8-2.0, which is the acceptable limit of contamination with proteins, and $A_{260/230}$: > 2,0 for contamination with carbohydrates. These data have also been reported by Sambrook et al. (1989) and Romano (1998). Furthermore, to avoid contamination with genomic DNA, Digest purified RNA with DNase I is needed to remove contaminating genomic DNA, which can act as template during PCR and may lead to spurious results. It may also be necessary to perform PCR on the treated RNA, using gene-specific primers, to confirm absence of genomic DNA (Udvardi et al., 2008).

To complete a reliability of the extracted and purified RNA, the integrity of the RNA requires evaluation. The measure of RNA reliability is based on the integrity of 28S and 18S ribosomal RNA and the lack thereof shows a smear in the agarose gel, indicating that the total RNA is degraded. Thus, an electrophoresis in agarose gel at 0.8% to 1.0% is recommended to observe the integrity of the ribosomal RNA bands (Sambrook et al., 1989).

5.3 Optimization and Efficiency Curve of Primers

Other parameters such as the optimization of primer concentrations and efficiency primer curves that might be done in serial dilutions or standard curves are important to perform qPCR assays. A dilution series of known template concentrations can be used to establish a standard curve for determining the initial starting amount of the target template or for assessing the reaction efficiency. The log of each known concentration in the dilution series is plotted against the Cq value for that concentration. Information on the performance of the reaction as well as various reaction parameters (including slope, y-intercept, and correlation coefficient) can be derived from this standard curve. The slope is obtained by the linear equation of the graph constructed by plotting the Cq values on the y-axis and the log values of the dilutions on the x-axis. The concentrations chosen for the standard curve should encompass the expected concentration range of the target (Pfaffl et al., 2004). At the end of the qPCR assay, primer efficiency must be calculated, and the formula most frequently described in the literature for this purpose is as follows: Efficiency = $10^{(-1/\text{slope})}$-1, in which the slope corresponds to the C_q value of the first dilution (concentration dilution) minus the C_q value of the last dilution divided by the number of dilutions. Hence, if the PCR is 100% efficient, the amount of PCR product will double with each cycle and the slope of the standard curve will be –3.33 (100 = 100% = $10^{(-1/-3.33)}$-1). The ideal slope is approximately -3.33 cycles; however, a slope between –3.9 and –3.0 (80-110% efficiency) is generally acceptable (Livak et al., 2001; Pfaffl et al., 2004). Calculated levels of target input may not be accurate if the reaction is not efficient. In order to improve efficiency, one must consider either (1) optimize primer concentrations or (2) design alternative primers.

Since SYBR Green binding dye is a non-specific dye that will detect any double-stranded DNA, it is important to verify if the qPCR is producing only the desired product. This can often be detected when PCR efficiencies are larger than 120% (Bustin et al., 2002; Bustin et al., 2009). Melting or dissociation curve is expressed during the last step of qPCR, following 40 cycles that only show one peak, revealing that a single multigene family isoform was amplified. These analyses can also be used to determine the approximate product size (Udvardi et al., 2004). If the melting curve has more than one major peak, the identities of the products should be determined by fractionating them on an ethidium DNA agarose gel electrophoresis to check for the presence of non-specific annealing. It must also be mentioned that lowering the primer concentrations will often reduce the amount of non-specific products. If the use of low primer levels still allow the detection of non-specific products in significant amounts, primer redesign may be a necessary measure. Once all cycles have been completed, the melting curve is added to evaluate the specificity of the primers. Melting curves with peaks lower than 78°C could indicate the presence of primer dimmers in the reaction or alternatively smaller non-specific amplicon products (Condori et al., 2001).

5.4 Normalisation and Analysis of qPCR Results

In gene-expression profile quantification, an assessment of the reliability of qPCR assay is required to normalize the target gene expression data. One of the most frequently used methods is the utilization of reference genes. Previous to the qPCR assay, it is necessary to design primers that amplify constitutive genes. The groups of reference genes are checked for stability to identify the most stable reference genes among all the selected genes that will be used to normalize the qPCR data, using programs such as geNorm, BestKeeper and NormFinder. Once the best reference genes are identified, data normalisation is required to ensure gene expression reliability (Zhong et al., 2011). Likewise, to confirm reliability of the results, biological and technical replicates must be obtained to provide data statistics, and evaluate the significance levels of gene expression analysis (Udvardi et al., 2004).

Relative quantification describes a real-time PCR experiment in which the gene of interest in one sample (i.e., treated) is compared to the same gene in another sample (i.e., untreated). The results are expressed as fold up- or down-regulation of the treated in relation to the untreated gene. Reference genes such as β-actin, GAPDH, elongation factor, among others are used as a control for experimental variability in this type of quantification (Tong et al., 2009). The most frequently used method for relative mRNA quantification by real time PCR has been described by Livak et al. (2001). This is a convenient method which presents the advantage of eliminating the need for standard curves. Thus, mathematical equations are used to calculate the relative expression levels of target relative to reference control or calibration, such as an untreated sample or RNA from normal tissue or a sample at time zero at qPCR experiments in time-course study. The amount of target gene in the sample normalized to a reference gene, relative to the normalized calibrator, is then given: $2^{-\Delta\Delta Cq}$, where $\Delta\Delta C_q = \Delta C_q$ (sample)- ΔCq (calibrator), and ΔC_q is the C_q of the target gene subtracted from the reference gene C_q, as describe by (Livak et al., 2001; Schmittgen et al., 2000; Schmittgen et al., 2008). In order to obtain reliable results, the target and reference gene must be approximately equal, and preferably at a percentage greater than 90%. This level of sequence equality is necessary to plot an efficiency curve based on the dilution serial method to given suitable results in experimental data, as described above. Finally, statistics methods, including student t-test, ANOVA, among others, must be applied to the concluding analysis. However, this method of Livak et al (2001) is limited due the use of only one reference gene. Actually, has been used more than one reference genes to normalisation data qPCR using algorithm that is based in the normalisation factor (NF) method, as geNorm, BestKeeper and NormFinder (Vandesompele et al., 2002; Andersen & Orntoft, 2004; Pfaffl et al., 2004) as described above.

6. Several Advances on the Real Time PCR

Several researchers are developing techniques to improve the quality of detection of DNA fluorescence. Recently, the manufacturer ELITE MGB™ has done a revolutionary advance in qPCR chemistry. The principle is based in the proprietary of the protein called Minor Groove Binder (MGB), Superbases™ and Eclipse®Dark Quencher technologies. These overlapping probes are much efficiently and accurately detect target DNA sequences, while offering greater sensitivity and specificity. According to manufacturer, the MGB protein is a synthetic molecule that binds to the minor groove of double stranded DNA molecules. In qPCR applications, MGB increases the stability of double stranded DNA complexes, specifically, the hybridization between the probe and the amplified DNA target. The increased DNA-DNA hybrid stability allows the design of shorter detection probes with higher specificity. Furthermore, The Eclipse®Dark Quencher is a proprietary fluorophore and dye quencher chemistry resulting in low background signals. Its key benefit is to ensure that every ELITe MGB™ assay will have the highest sensitivity by minimizing background signal interference. Together, show Real-Time PCR results of high accuracy.

Other works have shown important improving in the qPCR. Zheng et al. (2011) designed an aptamer-based sensing platform using a triple-helix molecular switch (THMS). The THMS consists of a central, target-specific aptamer sequence flanked by two arm segments and a dual-labeled oligonucleotide serving as a signal-transduction probe (STP). The STP is doubly labeled with pyrene at both ends and designed as a hairpin-shaped structure. Initially, the loop sequence of the STP binds with two arm segments of the aptamer, which forces the STP to form an ‘‘open’’ configuration and separate the two end labeled pyrene molecules, thus only emitting monomer fluorescence signal. The formation of the aptamer/target complex releases the STP, which switches to a ‘‘closed’’ hairpin configuration, bringing two pyrene molecules in close proximity and emitting excimer fluorescence signal. Hu et al. (2014) developed a modified Molecular Beacons–based multiplex qPCR Assay. In their work, all sets of primers and probes were combined, and the concentration of each reagent including primers, probes, magnesium, and Taq polymerase concentrations in the reaction mix were optimized. These modifications helped the sensitivity and specificity of the qPCR multiplex that were 100% and 99%, respectively.

Any need for fast and precise measurement of small amounts of nucleic acids represents a potential future niche for real-time PCR-based innovations. As machines become faster, cheaper, smaller, and easier to use through competition, standardized assay development, and advances in microfluidics (Mitchell et al., 2001), optics, and thermocycling, more in-field application needs are likely to be filled. In the commercial food industry and agriculture, real-time PCR will likely see expanded use for the detection and identification of microbes, parasites, or genetically modified organisms. Forensics will benefit from real-time PCR's sensitivity, specificity, and speed, especially because time is crucial to many criminal investigations and specimen size may be limited. Reduced cost and increased portability open the door for the diagnosis of diseases in remote areas along with on-site epidemiological studies and may facilitate the transfer of needed scientific technologies to developing countries, thereby contributing to their “scientific capacity”.

7. Final Considerations

The polymerase chain reaction (PCR) is one of the most powerful technologies in molecular biology. qPCR is an efficient tool to measure the levels of mRNA expression in different types of samples; their use together with the reference genes are ideal for decreasing the possible errors in RNA extraction and contamination during sample manipulation, thus increasing the quality of cDNA. The qPCR has the sensible technical power to amplify target specific genes, but it is necessary to obtain reliable results in the gene expression profile. Several parameters must be considered, including good design of primers, evaluating their specificity and efficiency. In addition, the RNA extracted must be free of contaminants, such as carbohydrates, proteins and phenols, because these may interfere with the polymerases during PCR reaction. For the normalisation of qPCR data, the use of reference gene is needed to provide suitable results and reproducibility. Thus, the selection of a reference gene for each experimental condition is crucial. These precautions are pivotal to render reliable results during gene expression analysis.

Acknowledgements

The authors are grateful to CNPq and CAPES. This work was supported by the Universidade Federal do Ceará-UFC.

References

Andersen, C. L., Jensen, J. L., & Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer research*, *64*(15), 5245-5250. http://dx.doi.org/10.1158/0008-5472.CAN-04-0496

Barsalobres-Cavallari, C. F., Severino, F. E., Maluf, M. P., & Maia, I. G. (2009). Identification of suitable internal control genes for expression studies in Coffea arabica under different experimental conditions. *BMC molecular biology*, *10*(1), 1. http://dx.doi.org/10.1186/1471-2199-10-1

Bustin, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of molecular endocrinology*, *29*(1), 23-39. http://dx.doi.org/10.1677/jme.0.0290023

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., ... & Wittwer, C. T. (2009). The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry*, *55*(4), 611-622. http://dx.doi.org/10.1373/clinchem.2008.112797

Condori, J., Nopo-Olazabal, C., Medrano, G., & Medina-Bolivar, F. (2011). Selection of reference genes for qPCR in hairy root cultures of peanut. *BMC research notes*, *4*(1), 392. http://dx.doi.org/10.1186/1756-0500-4-392

Cordoba, E. M., Die, J. V., González-Verdejo, C. I., Nadal, S., & Román, B. (2011). Selection of reference genes in Hedysarum coronarium under various stresses and stages of development. *Analytical biochemistry*, *409*(2), 236-243. http://dx.doi.org/10.1016/j.ab.2010.10.031

De Boever, S., Vangestel, C., De Backer, P., Croubels, S., & Sys, S. U. (2008). Identification and validation of housekeeping genes as internal control for gene expression in an intravenous LPS inflammation model in chickens. *Veterinary immunology and immunopathology*, *122*(3), 312-317. http://dx.doi.org/10.1016/j.vetimm.2007.12.002

Dekkers, B. J., Willems, L., Bassel, G. W., van Bolderen-Veldkamp, R. M., Ligterink, W., Hilhorst, H. W., & Bentsink, L. (2012). Identification of reference genes for RT–qPCR expression analysis in Arabidopsis and tomato seeds. *Plant and Cell Physiology*, *53*(1), 28-37. http://dx.doi.org/10.1093/pcp/pcr113

Demidenko, N. V., Logacheva, M. D., & Penin, A. A. (2011). Selection and validation of reference genes for quantitative real-time PCR in buckwheat (Fagopyrum esculentum) based on transcriptome sequence data. *PLoS One*, *6*(5), e19434. http://dx.doi.org/10.1371/journal.pone.0019434

Dheda, K., Huggett, J. F., Bustin, S. A., Johnson, M. A., Rook, G., & Zumla, A. (2004). Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *Biotechniques*, *37*, 112-119.

Expósito-Rodríguez, M., Borges, A. A., Borges-Pérez, A., & Pérez, J. A. (2008). Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biology*, *8*(1), 131. http://dx.doi.org/10.1186/1471-2229-8-131

Foss, D. L., Baarsch, M. J., & Murtaugh, M. P. (1998). Regulation of hypoxanthine phosphoribosyltransferase, glyceraldehyde‐3‐phosphate dehydrogenase and β‐actin mRNA expression in porcine immune cells and tissues. *Animal biotechnology*, *9*(1), 67-78. http://dx.doi.org/10.1080/10495399809525893

Hamalainen, H. K., Tubman, J. C., Vikman, S., Kyrölä, T., Ylikoski, E., Warrington, J. A., & Lahesmaa, R. (2001). Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT-PCR. *Analytical biochemistry*, *299*(1), 63-70. http://dx.doi.org/10.1006/abio.2001.5369

Holland, M. J. (2002). Transcript abundance in yeast varies over six orders of magnitude. *Journal of Biological Chemistry*, *277*(17), 14363-14366. http://dx.doi.org/10.1074/jbc.C200101200

Hong, S. Y., Seo, P. J., Yang, M. S., Xiang, F., & Park, C. M. (2008). Exploring valid reference genes for gene expression studies in Brachypodium distachyon by real-time PCR. *BMC plant biology*, *8*(1), 112. http://dx.doi.org/10.1186/1471-2229-8-112

Hwang, S., Kang, B., Hong, J., Kim, A., Kim, H., Kim, K., & Cheon, D. S. (2013). Development of duplex real‐time RT‐PCR based on Taqman technology for detecting simultaneously the genome of pan‐enterovirus and enterovirus 71. *Journal of medical virology*, *85*(7), 1274-1279. http://dx.doi.org/10.1002/jmv.23588

Jain, M., Nijhawan, A., Tyagi, A. K., & Khurana, J. P. (2006). Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and biophysical research communications*, *345*(2), 646-651. http://dx.doi.org/10.1016/j.bbrc.2006.04.140

Kumar, K., Muthamilarasan, M., & Prasad, M. (2013). Reference genes for quantitative real-time PCR analysis in the model plant foxtail millet (Setaria italica L.) subjected to abiotic stress conditions. *Plant Cell, Tissue and Organ Culture (PCTOC)*, *115*(1), 13-22. http://dx.doi.org/10.1007/s11240-013-0335-x

La Cruz, S., Lopez-Calleja, M. I., Alcocer, M., González, I., Martín, R., & García, T. (2013). TaqMan real-time PCR assay for detection of traces of Brazil nut (*Bertholletia excelsa*) in food products. *Food control,* 140, 382-389. http://dx.doi.org/10.1016/j.foodcont.2013.01.053

Lee, J. M., Roche, J. R., Donaghy, D. J., Thrush, A., & Sathish, P. (2010). Validation of reference genes for quantitative RT-PCR studies of gene expression in perennial ryegrass (Lolium perenne L.). *BMC Molecular Biology*, *11*(1), 8. http://dx.doi.org/10.1186/1471-2199-11-8

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *methods*, *25*(4), 402-408. http://dx.doi.org/10.1006/meth.2001.1262

Mallona, I., Lischewski, S., Weiss, J., Hause, B., & Egea-Cortines, M. (2010). Validation of reference genes for quantitative real-time PCR during leaf and flower development in Petunia hybrida. *BMC Plant Biology*, *10*(1), 4. http://dx.doi.org/10.1186/1471-2229-10-4

Marshall, O. J. (2004). PerlPrimer: cross-platform, graphical primer design for standard, bisulphite and real-time PCR. *Bioinformatics 20*(15): 2471–2472. http://dx.doi.org/10.1093/bioinformatics/bth254

Mitchell, P. (2001). Microfluidics-downsizing large-scale biology. Nature biotechnology, 19(8), 717-721. http://dx.doi.org/10.1093/bioinformatics/bth254

Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in enzymology*, *155*, 335-350. http://dx.doi.org/10.1016/0076-6879(87)55023-6

Ng, C. T., Gilchrist, C. A., Lane, A., Roy, S., Haque, R., & Houpt, E. R. (2005). Multiplex real-time PCR assay using Scorpion probes and DNA capture for genotype-specific detection of Giardia lamblia on fecal samples. *Journal of clinical microbiology*, *43*(3), 1256-1260. http://dx.doi.org/10.1128/JCM.43.3.1256-1260.2005

Niu, J. Z., Dou, W., Ding, T. B., Yang, L. H., Shen, G. M., & Wang, J. J. (2012). Evaluation of suitable reference genes for quantitative RT-PCR during development and abiotic stress in Panonychus citri (McGregor)(Acari: Tetranychidae). *Molecular biology reports*, *39*(5), 5841-5849. http://dx.doi.org/10.1007/s11033-011-1394-x

Pabla, S. S., & Pabla, S. S. (2008). Real-time polymerase chain reaction. *Resonance*, *13*(4), 369-377. http://dx.doi.org/10.1007/s12045-008-0017-x

Petit, C., Pernin, F., Heydel, J. M., & Délye, C. (2012). Validation of a set of reference genes to study response to herbicide stress in grasses. *BMC research notes*, *5*(1), 18. http://dx.doi.org/10.1186/1756-0500-5-18

Pfaffl, M. W., Tichopad, A., Prgomet, C., & Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. *Biotechnology letters*, *26*(6), 509-515. http://dx.doi.org/10.1023/B:BILE.0000019559.84305.47

Ransbotyn, V., & Reusch, T. B. (2006). Housekeeping gene selection for quantitative real‐time PCR assays in the seagrass Zostera marina subjected to heat stress. *Limnology and Oceanography: Methods*, *4*(10), 367-373. http://dx.doi.org/10.4319/lom.2006.4.367

Rebouças, E. D. L., Costa, J. J. D. N., Passos, M. J., Passos, J. R. D. S., Hurk, R. V. D., & Silva, J. R. V. (2013). Real time PCR and importance of housekeepings genes for normalization and quantification of mRNA expression in different tissues. *Brazilian Archives of Biology and Technology*, *56*(1), 143-154. http://dx.doi.org/10.1590/S1516-89132013000100019

Reid, K. E., Olsson, N., Schlosser, J., Peng, F., & Lund, S. T. (2006). An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC plant biology*, *6*(1), 27. http://dx.doi.org/10.1186/1471-2229-6-27

Rocha, A. J., Miranda, R., & Cunha, R. M. S. (2014). Avaliação de DNA polimerases em ensaios de amplificação de microssatélites através do PowerPlex® 16 BIO System. *BBR-Biochemistry and Biotechnology Reports*, *3*(2), 1-8. http://dx.doi.org/10.5433/2316-5200.2014v3n2p1

Rocha, A. J., Soares, E. L., Costa, J. H., Costa, W. L., Soares, A. A., Nogueira, F. C., ... & Campos, F. A. (2013). Differential expression of cysteine peptidase genes in the inner integument and endosperm of developing seeds of Jatropha curcas L.(Euphorbiaceae). *Plant science*, *213*, 30-37. http://dx.doi.org/10.5433/2316-5200.2014v3n2p1

Romano, E., Brasileiro, A. C. M. (1998). Extração de DNA de Tecidos Vegetais. In A. C. M. Brasileiro, & V. T. C. (Eds.), Carneiro *Manual de transformações Genéticas De Plantas***.** Editora Embrapa: Brasília v. 40-43, 1998.

Saha, G. C., & Vandemark, G. J. (2013). Stability of expression of reference genes among different Lentil (Lens culinaris) genotypes subjected to cold stress, white mold disease, and aphanomyces root rot. *Plant Molecular Biology Reporter*, *31*(5), 1109-1115. http://dx.doi.org/10.1007/s11105-013-0579-y

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, *230*(4732), 1350-1354. http://dx.doi.org/10.1126/science.2999980

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning* (Vol. 2, pp. 14-9). New York: Cold spring harbor laboratory press.

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature protocols*, *3*(6), 1101-1108. http://dx.doi.org/10.1038/nprot.2008.73

Schmittgen, T. D., & Zakrajsek, B. A. (2000). Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *Journal of biochemical and biophysical methods*, *46*(1), 69-81. http://dx.doi.org/10.1016/S0165-022X(00)00129-9

Selim, M., Legay, S., Berkelmann-Löhnertz, B., Langen, G., Kogel, K. H., & Evers, D. (2012). Identification of suitable reference genes for real-time RT-PCR normalization in the grapevine-downy mildew pathosystem. *Plant cell reports*, *31*(1), 205-216. http://dx.doi.org/10.1007/s00299-011-1156-1

Suzuki, T., Higgins, P. J., & Crawford, D. R. (2000). Control selection for RNA quantitation. *Biotechniques*, *29*(2), 332-337.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., ... & Heinen, E. (1999). Housekeeping genes as internal standards: use and limits. *Journal of biotechnology*, *75*(2), 291-295. http://dx.doi.org/10.1016/S0168-1656(99)00163-7

Tong, Z., Gao, Z., Wang, F., Zhou, J., & Zhang, Z. (2009). Selection of reliable reference genes for gene expression studies in peach using real-time PCR. *BMC Molecular Biology*, *10*(1), 71. http://dx.doi.org/10.1186/1471-2199-10-71

Udvardi, M. K., Czechowski, T., & Scheible, W. R. (2008). Eleven golden rules of quantitative RT-PCR. *The Plant Cell Online*, *20*(7), 1736-1737. http://dx.doi.org/10.1105/tpc.108.061143

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, *3*(7), research0034. http://dx.doi.org/10.1186/gb-2002-3-7-research0034

VanGuilder, H. D., Vrana, K. E., & Freeman, W. M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques*, *44*(5), 619-626. http://dx.doi.org/10.2144/000112776

Wong, M. L., & Medrano, J. F. (2005). Real-time PCR for mRNA quantitation. *Biotechniques*, *39*(1), 75-85. http://dx.doi.org/10.2144/05391RV01

Zheng, J., Li, J., Jiang, Y., Jin, J., Wang, K., Yang, R., & Tan, W. (2011). Design of aptamer-based sensing platform using triple-helix molecular switch. *Analytical chemistry*, *83*(17), 6586-6592. http://dx.doi.org/10.2144/05391RV01

Zhong, H. Y., Chen, J. W., Li, C. Q., Chen, L., Wu, J. Y., Chen, J. Y., ... & Li, J. G. (2011). Selection of reliable reference genes for expression studies by reverse transcription quantitative real-time PCR in litchi under different experimental conditions. *Plant cell reports*, *30*(4), 641-653. http://dx.doi.org/10.1007/s00299-010-0992-8

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

Study on the Heterosis of the First Generation of Hybrid between Chinese and Korean Populations of S*capharca broughtonii* using Methylation-Sensitive Amplification Polymorphism (MSAP)

Hailin Sun[1], Yanxin Zheng[2], Chunnuan Zhao[2], Tao Yu[2] & Jianguo Lin[2]

[1] Changdao Fisheries Research Institute, Changdao, China

[2] Changdao Enhancement and Experiment Station, Chinese Academy of Fishery Sciences, Changdao, China

Correspondence: Tao Yu, Changdao Enhancement and Experiment Station, Chinese Academy of Fishery Sciences, Changdao, China. Email: cdyutao@126.com

Received: September 9, 2015 Accepted: September 20, 2015 Online Published: November 6, 2015

doi:10.5539/jmbr.v5n1p56 URL: http://dx.doi.org/10.5539/jmbr.v5n1p56

Abstract

DNA methylation is known to play an important role in the regulation of gene expression in eukaryotes. In this study, the author assessed the extent and pattern of cytosine methylation in the *Scapharca broughtonii* genome using the technique of methylation-sensitive amplified polymorphism (MSAP).The results showed that, DNA methylation rate was negatively related to the shell length, the gross weight and the weight of soft body, but positively related to the shell broadness and the shell height; there was significantly different between the parents and the offspring: 31.6% of 5'-CCGG sites in the *Patinopecten yessoensis* of Korean populations genome were cytosine methylated, and in the *Patinopecten yessoensis* of Chinese populations were 33%, the methylation rates of F1 was 29.98%; four classes of patterns were identified in a comparative assay of cytosine methylation in the parents and hybrid, increased methylation was detected in the hybrid compared to the parents at some of the recognition sites, while decreased methylation in the hybrid was detected at other sites. It indicated that the alteration of methylation resulted from cross-breeding, and the inbreeding did not change the methylation ratio and patterns; The DNA cytosine methylation has a relationship with the heterosis.

Keywords: *Scapharca broughtonii*, DNA methylation, methylation-sensitive amplified polymorphism (MSAP), heterosis

1. Introduction

Scapharca broughtonii (Mollusca, Bivalve, Arcoida), one of the most important marine commerical bivalve species, mainly distributes in the coasts of BoHai Sea and North of Yellow Sea, China. Because of its large body, rapid growth rate, delicious tastes and high protein and vitamin contents, the export in the exchange rates of ark shell is higher in the aquatic products, and *Scapharca broughtonii* has become one of the most popular farming mollusks in North China due to its high economic value in recent years. However, with the deterioration of marine ecological environment resulted from the extended farming scale and frequency human activities in coastal waters, and over-fishing, the wild resources decreased. The mass mortality has become a major constraint for the development of the *Scapharca broughtonii* culture. It is imperative for us to actively manage the resource and turn to breed high adversity resistance, fast-growing variety using traditional and new breeding methods. The hybridization of different populations proved to be a good way of breeding.

The genetic basis of heterosis has been debated for decades, dominance, pseudo-overdominance, real overdominance, and epistasis are the major genetic models proposed to explain heterosis (Crow, 2000; Lamkey & Edwards, 1997; Lippman & Zamir, 2007; Reif et al., 2006), but there is still a striking discordance between an extensive use of heterosis in variety development and our understanding of the basis of heterosis (Birchler, Auger, & Riddle, 2003; Reif et al., 2006). In recent years, many research thought that the molecular basis of heterosis may be attributed to the increased gene expression level in the hybrid or to the altered regulation of gene expression in the hybrid either at the global level or for specific classes of genes (Leonardi, Damerval, Hebert, Gallais, & Vienne, 1991; Romagnoli, Maddaloni, Livini, & Motto, 1990; A. Tsaftaris, Kafka, Polidoros, & Tani, 1997; S. Tsaftaris, 2006). Two different alleles brought together in the hybrid may create a combined allelic expression pattern in the hybrids. Alternatively, at some loci, allelic interaction or a change in the spectrum of trans-acting

factors causes gene expression in the hybrid to deviate from simple additive allelic expression patterns of the parents(Birchler et al., 2003; Gibson & Weir, 2005). Considering effects of DNA methylation on gene expression, there may be a relationship between DNA methylation and the expression of heerosis (Finnegan, Peacock, & Dennis, 2000; Rangwala & Richards, 2004).

DNA cytosine methylation is the most common covalent modification of DNA in eukaryotes, in recent years, DNA methylation has received considerable attention in eukaryotic organisms (Xiong, Xu, Saghai Maroof, & Zhang, 1999), which plays an important role in many aspects of biology, including differential gene expression, cell differentiation, genomic imprinting, chromatin inactivation, transposable elements and gene silencing, and so on(Finnegan et al., 2000; Paszkowski & Whitham, 2001; Tariq & Paszkowski, 2004).

DNA methylation analysis has been approached either by studying global levels of cytosines methylated or by analyzing specific gene sequences(Jacobsen, Sakai, Finnegan, Cao, & Meyerowitz, 2000; Luff, Pawlowski, & Bender, 1999; Riddle & Richards, 2002; Soppe et al., 2000). There are several methods used for detecting DNA methylation, such as bisulfite conversion, methylation-sensitive restriction enzymes, methyl-binding proteins, methylation-sensitive amplified polymorphism (MSAP), and anti -methylation cytosine antibodies(Zilberman & Henikoff, 2007). Among these, two methods are routinely used for the detection of DNA methylation in the tissues of eukaryotic organisms. These depend on the application of bisulfites or methylation-sensitive restriction enzymes. Bisulfites convert unmethylated cytosine into thymine, thus allowing the detection of cytosine methylation. Some restriction enzymes (isoschizomers) share the same recognition sites but show differential sensitivity to DNA methylation. Thus, polymorphic DNA fragments can be generated after digestion of methylated genomic DNA with isoschizomers(Xu, Li, & Korban, 2000). Methylation sensitive amplified polymorphism (MSAP) analysis is based on the use of isoschizomers for detection of DNA methylation. It is an adaptation of the amplified fragment length polymorphism (AFLP) technique(Reyna-Lopez, Simpson, & Ruiz-Herrera, 1997), in which the isoschizomers HpaII and MspI are employed as 'frequent-cutter' enzymes for AFLP, instead of the usual MseI. HpaII and MspI recognize the same tetranucleotide sequence (5'-CCGG-3'), but display differential sensitivity to DNA methylation. HpaII is inactive when either of the two cytosines is fully methylated, but cleaves hemi-methylated 5'-CCGG-3' at a lower rate than the unmethylated sequence. MspI cleaves 5'-C^{5m}CGG-3', but not 5'-^{5m}CCGG-3'. MSAP allows for detection of genetic diversity throughout the genome without any prior knowledge of the nucleotide sequence(Vos et al., 1995) and has been successfully applied in various studies.

In this study, the MSAP technique was firstly used to analyze the *Scapharca broughtonii* genome DNA cytosine methylation. We discussed the differences in the level of cytosine methylation among the parents and the offspring, the differences of methylation patterns between the parents and offspring, the correlation between the methylation and phenotypic traits, and emphasized the discussion on the molecular basis of heterosis in terms of the DNA methylation.

2. Materials and Methods

2.1 Sampling

The *Scapharca broughtonii*, Chinese populations(6.9±0.53 cm of shell length), were collected randomly from Penglai sea area (Shandong province, China) in April, 2014. Korean populations (7.2±0.47 cm)were collected randomly from Incheon sea area in April，2014. The heterozygous **F1** (6.7±0.62 cm) was a cross between the *Scapharca broughtonii* of Chinese and Korean populations**.**

2.2 DNA Extraction

The sample DNA was extracted from adductor muscle by traditional phenol-chloroform method. Approximately 100 mg of adductor muscle was dissected out and transferred to an Eppendorf tube containing 500μL of a lysis solution (50 mmolL-1 Tris–Cl pH8.0, 10 mmolL-1 EDTA, 10% sodium dodecylsulfate[SDS] and 200 μgmL-1 proteinase K) at 55 ℃ for 3 h. DNA was extracted with phenol, phenol/chloroform/isoamyl alcohol (25:24:1), and chloroform, and then precipitated with two volumes of ice-chilled absolute ethanol and 1/10 volume of 3 mol L-1 sodium chloride at -20 ℃ for 1h. The rough extraction was washed with 70% ethanol for three times, natural dried and resuspended in 50μL autoclaved ddH_2O. Extracted DNA was stored at-20 ℃.

2.3 MSAP Analysis

The MSAP protocol was adapted from Xiong et al(Xiong et al., 1999). Briefly, DNA was double-digested with one of the methylation sensitive enzymes HpaII or MspI, which cuts at the CCGG site, and then with the methylation insensitive EcoRI. Two digestion reactions were set up at the same time for each genomic DNA sample, each containing 400 ng of DNA with 3 Units of isoschizomers either HpaII or MspI (Fermentas) and 2μL

10×Buffer Tango™ in a final volume of 20 μL for 6 h at 37°C, and then add in 3 Units of EcoRI (Fermentas) and 4μL 10×Buffer Tango™ in a final volume of 30 μL for 6 h at 37 ℃.

Subsequently, the digested DNA fragments from the two reactions were ligated separately with an equal volume of the ligation solution containing 5μL digested fragments with 5 U of T4 DNA Ligase (Trans), 5 pmolL-1 EcoRI adapter, 50 pmolL-1 HpaII/MspI adapter, and 4μL 5×T4 DNA Ligase Buffer in a final volume of 20 μL at 16℃ for overnight. The reactions were stopped by incubating at 65°C for 10 min and diluted to 200μL for PCR amplification.

Preamplification was conducted by using 5μL of the above ligation product with E0/HM0 primers in a volume of 20μL containing 2 μL PCR buffer, 0.1 mmolL-1 each dNTPs, 20 ng of each primer and 0.1 U Taq polymerase (TIANGEN). The reaction involved 27 cycles of 94°C for 30 s, 56°C for 1min, 72°C for 1 min, with a final extension at 72°C for 10 min. The preamplified products were then diluted to 600μL and stored at -20°C before use.

Ingredients of the selective amplification were the same as described above using 2.5μL of diluted preamplification mixture DNA. The selective amplification was performed by the touchdown program using amplification primers. The PCR conditions were as follows: 13 cycles at 94℃ for 30 s, 0.7℃ per cycle from 65 to 56℃ for 30 s and 72℃ for 1 min; and another 23 cycles of PCR amplification were used following the touchdown program. The denaturing step was done at 94℃ for 30 s, annealing at 56℃ for 30 s, extension at 72℃ for 1 min; and a final extension at 72℃ for 10 min.

The final selective amplification products were denatured, separated on a 6% polyacrylamide sequencing gel, and visualized by silver staining.

All reactions were performed in triplicate to avoid false positive results. If results were reproducible, the sample was used for further analysis. Only clear and reproducible bands that appeared in four independent PCR amplifications were scored.

2.4 Restriction of Isolated and Re-Amplified Fragments

The special bands were excised directly from the polyacrylamide gels on the plate using a razor blade. The bands were rehydrated with 50μL of sterile distilled water, heated at 98℃ for 5 min and let cool slowly to room temperature for the night. The tubes were centrifuged at 12,000 g for 10 min and the supernatant transferred into a fresh tube. Aliquots of 5μL were used as template for re-amplification in a total PCR reaction volume as the selective amplification with the same primer combinations. The products were checked on 1.5% agarose gel for the presence of the bands.

Then, two sets of digestion reactions were carried out simultaneously, in the first reaction, 5μL of re-amplification PCR product was added to 15μL of the digestion system above described, the reamplified fragment was excised from the H lane. The second digestion reaction was carried out in the same way, except that MspI was used in place of HpaII, and the reamplified fragment was excised from the M lane. The digestion products were checked on 1.5% agarose gel with the re-amplification PCR product.

3. Results

The isoschizomers HpaII and MspI recognize and digest 5'-CCGG-3 sites, but display differential sensitivity to DNA methylation. HpaII is inactive if one or both cytosines are fully methylated (both strands methylated), but cleaves hemimethylated sequences (only a single DNA strand is methylated) or no methylation sequences; whereas, MspI digests inner methylation of double-stranded DNA or no methylation. Hemimethylation of either of the two cytosines would lead to the appearance of a fragment in the amplification product from the EcoRI+HpaII digest but not the EcoRI+MspI digest; on the contrary, it is a full methylation site; and if the fragments appeared in the products of the two digestions, the cytosine were not methylated(Lu et al., 2006).

3.1 The Correlation between Methylation Rates and Phenotypic Traits

The methylation may affected *Scapharca broughtonii* phenotypic traits were studied with adductor muscle DNA samples of *Scapharca broughtonii* using 9 pairs of primers (Table 1). And Figure 1 show the correlation between methylation rates and phenotypic traits, results showed that DNA methylation rate was negatively related to the shell length, the gross weight and the weight of soft body, but positively related to the shell broadness and the shell height, there was great correlativity between the DNA methylation and the gross weight (from Korean, the correlation coefficient is -0.59) (Figure 1). It indicated that DNA methylation affected on the shell length, the gross weight and the weight of soft body, DNA methylation may play an important part during the growth and development of the organisms, in the aspect of survival rate it has a different effect on Chinese and Korean populations.

Table 1. List of MSAP primers and adapters used

	EcoRI	MspI/HpaII
Adapters	EA_1:5'-CTC GTA GAC TGC GTA CC-3' EA_2:5'-AAT TGG TAC GCA GTC TAC-3'	HMA_1:5'-GAT CAT GAG TCC TGC T-3' HMA_2:5'-CGA GCA GGA CTC AGA A-3'
Primers for Preamplification	E_0:5'-GAC TGC GTA CCA ATT C-3'	HM_0:5'-ATC ATG AGT CCT GCT CGG G-3'
Primers for Selective Amplification	E_1: 5'-GAC TGC GTA CCA ATT C ACA-3' E_2: 5'-GAC TGC GTA CCA ATT C AGT-3' E_3: 5'-GAC TGC GTA CCA ATT C AAC-3' E_4: 5'-GAC TGC GTA CCA ATT C GTC-3' E_5: 5'-GAC TGC GTA CCA ATT C GCT-3'	HM_1: 5'-ATC ATG AGT CCT GCT CGG GC TGA-3' HM_2: 5'-ATC ATG AGT CCT GCT CGG GC TGT-3' HM_3: 5'-ATC ATG AGT CCT GCT CGG GC TAT-3' HM_4: 5'-ATC ATG AGT CCT GCT CGG GC TAC-3' HM_5: 5'-ATC ATG AGT CCT GCT CGG GC TCA-3' HM_6: 5'-ATC ATG AGT CCT GCT CGG GC TCT-3' HM_7: 5'-ATC ATG AGT CCT GCT CGG GC TTC-3' HM_8: 5'-ATC ATG AGT CCT GCT CGG GC TTA-3'

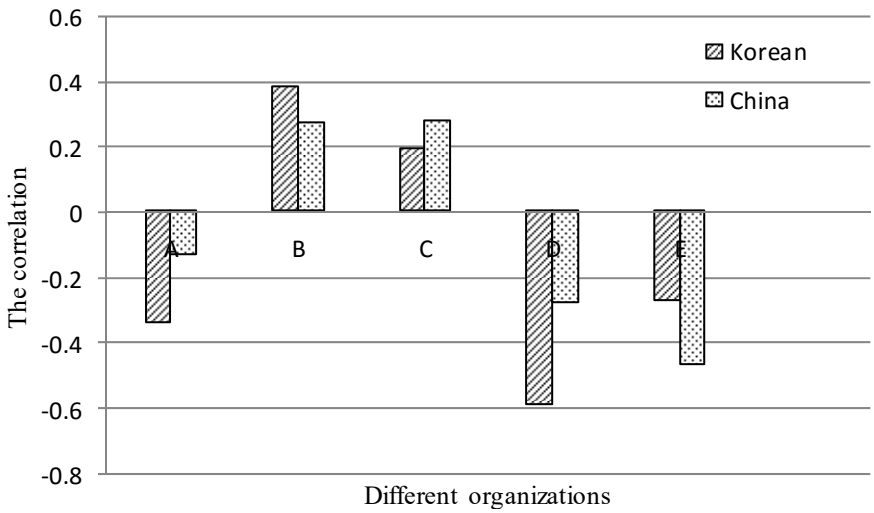

Figure 1. The correlation between methylation rates and phenotypic traits

A: The shell length; B: the shell broadness; C: the shell height; D: the gross weight; E: the weight of soft body

3.2 The Methylation Rates of Parental Lines and Hybrid

The two parental lines, their F1 hybrid were compared using tissue from adductor muscle with the same primers (Table 1). A total of 732 fragments were amplified, each of the fragments represented a recognition site cleaved by one or both of the isoschizomers. Firstly, the two parents showed significantly different degree of methylation (Figure 2): 197 differentially amplified fragments were detected in *Scapharca broughtonii* from Korean and 183 were observed in *Scapharca broughtonii* from China. Thus, approximately 31.6% of 5'-CCGG sites in *Scapharca broughtonii* (from Korean) genome were cytosine-methylated, in *Scapharca broughtonii* from China were 33%. In *Scapharca broughtonii* from Korean, the full methylation rate was 26.58%, the hemimethylation rate was 4.82%, in *Scapharca broughtonii* from China, the full methylation rate was 28.68%, and the hemimethylation rate was 5.39%.The methylation rates of F1was less than the parents (Figure 2), a total of 231 fragments were amplified, and the rates of the methylation were 29.98%. In F1, full methylation of internal cytosines accounted for 79.94% of the methylated sites, and the remaining 20.06% were due to hemimethylation. The F1 inclined to the female parent on the traits, so the methylation rate was close to the *Scapharca broughtonii* from Korean, it was 31.6%.

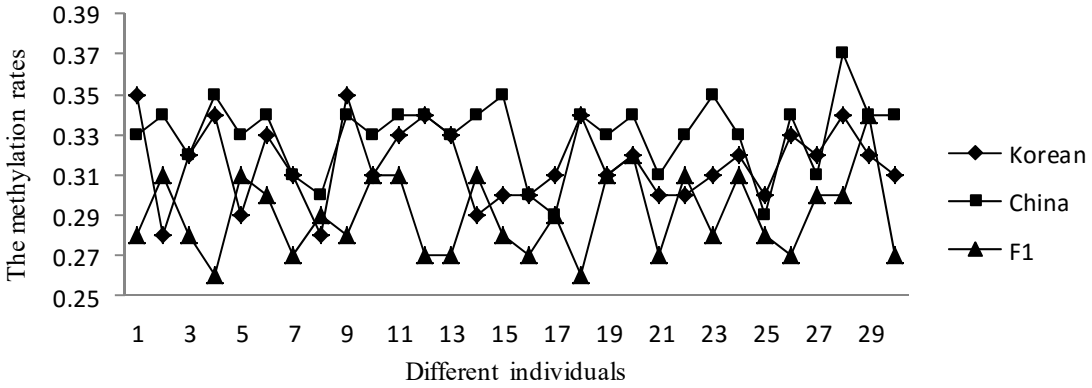

Figure 2. The methylation rates of parental lines and hybrid

Each group chose 30 individuals; each site stands for the average methylation rate of 9 pairs of primes.

3.3 Differential Methylation Patterns among Parental Lines and Hybrid

Four major classes of banding patterns were identified among the differentially amplified fragments (Table 3). In the first class (class A), the same methylation sites were detected in both parents and in the hybrid; these are referred to as monomorphic with respect to cytosine methylation, within the resolving power of this technique. In the four groups, 19 sites detected by 9 primer pairs reflected full methylation of the internal cytosine, and 6 sites were the result of hemimethylation. The second class (class B) showed simple Mendelian inheritance of the methylated bands，irrespective of the enzyme digest: A band that was detected in either parent was also detected in the hybrids. Class B could be divided in to four subclasses，B1,B2, B3 and B4 (Table 2). B1 averagely accounted for 3.1% methylated sites，B2 did 16.7%,B3 did 2.1% and B4 did 26.0% methylated sites. This class accounted for totally 46.9%methylated sites. Class C represents an increase in the level of methylation in the hybrid compared to the parental lines: a site detected in one or both parental lines was not observed in the hybrid (Table 2).Class C could be divided into 3 subclasses, C1 (a band was only revealed in *P. yessoensis*), C2 (a band was only revealed in *C. farreri*), C3 (a band was revealed in both of *C. farreri* and *P. yessoensis*). On the contrary, class D indicated demethylation in the hybrid genome. That is, a band was observed in the hybrids, but not in the parental lines (Table 2).The two classes occupied 26% methylated sites; with 18.7%belong to C and 7.3% to D.

Table 3. Patterns of cytosine methylation in parental lines and their F1 hybrid

Patterns	*Scapharca broughtonii* from Korean		*Scapharca broughtonii* from China		F1 hybrid		Number of sites
	HpaII	MspI	HpaII	MspI	HpaII	MspI	
A1	-	+	-	+	-	+	19
A2	+	-	+	-	+	-	6
B1	-	+	-	-	-	+	3
B2	-	-	-	+	-	+	16
B3	+	+	-	+	+	+	2
B4	-	+	+	+	+	+	25
C1	-	+	-	-	-	-	10
C2	-	-	-	+	-	-	6
C3	-	+	-	+	-	-	2
D1	-	+	-	-	+	+	0
D2	-	-	-	+	+	+	7

"-"stands for no band was found at the site; "+" stands for a band was found at the site.

4. Discussion

In recent years, there has been an increased interest in understanding the role of DNA methylation in regulating gene expression during the growth and development of the organisms (Chen, Ma, Chen, Song, & Zhang, 2009). The methylation sensitive amplified polymorphism (MSAP) technique has been used in various studies on cytosine methylation, and has proven to be a powerful tool for investigating DNA methylation. In this study, we have adapted the MSAP technique for the detection of cytosine methylation in the *Scapharca broughtonii* genome, the results showed that this technique is highly efficient for investigating the cytosine methylation of *Scapharca broughtonii*; all bands detected displayed good stability, reproducibility, and consistency.

The DNA methylation levels are varied in different species: 16.3% of the methylation has been reported to occur in the rice genome(Xiong et al., 1999), 35–43% in *Arabidopsis* of different ecological type(Cervera, Ruiz-Garcia, & Martinez-Zapater, 2002), 33% in wheat(Horváth et al., 2002), 15.7% in the developing seeds from Brassica napus(Lu et al., 2006). In animals, the pigs' methylation ratio is about 10%; the grass carp's methylation is very high, about 75.9%. Our research found that 31.6% of 5'-CCGG sites in *Scapharca broughtonii* from Korean genome were cytosine methylated, in *Scapharca broughtonii* from China were 33%, the DNA methylation level is different significantly to other organism, the differences may come from the detection method(such as the number of primes, the reaction conditions and time),the experiment material (different tissues: adductor muscle, mantle, gill filaments, gut, and gonad; or different collection time), and the genetic factors(the genetic factors play a more important part). Riddle (Riddle & Richards, 2002) found that natural variation in NOR methylation results from a combination of genetic and epigenetic mechanisms.

The purpose of this study was to develop an approach to investigating the possible role of methylation in the expression of heterosis. The genetic basis of heterosis has been debated for decades, but the mechanism underlying heterosis remains mysterious. Before the 1990s, two major hypotheses have been promulgated to explain this phenomenon: the dominance hypothesis and the overdominance hypothesis, but they are ideal and very limited, because they all thought that one trait was controlled by one allele. The fact is that more and more studies show that the appearance of quantitative character was the result of many alleles acting together. Many researchers used QTL method to study the heterosis, and put forth the hypothesis of epistasis, but the results were also not satisfactory (Z. Li et al., 2001; Luo et al., 2001; Xiao, Li, Yuan, & Tanksley, 1995; Yu et al., 1997).

Now, many researchers believe that the molecular basis of heterosis may be attributed to the increased gene expression level in the hybrid(Leonardi et al., 1991; Romagnoli et al., 1990; A. Tsaftaris et al., 1997; S. Tsaftaris, 2006) or to the altered regulation of gene expression in the hybrid either at the global level or for specific classes of genes.

The phenotypic traits that the organism expressed are the results of gene expression, an individual take on an advantage in a trait, it should be the result of over-expression of some genes or under-expression of some genes, that is to say, the over-expression and under-expression of some genes lead to the expression of heterosis together. Romagnoli et al. (Romagnoli et al., 1990) found that heterosis may derive from simple dominant or codominant gene effects in addition to the increased expression of certain loci. Tsaftaris et al.(A. Tsaftaris et al., 1997) found that the transcriptional level of 35 gene were higher than the parents lines. Sun et al.(Sun et al., 2004) found that 30% genes were differentially expressed between hybrids and their parents, which play an important role for hybrids to demonstrate heterosis. Li et al.(X. Li, Wei, Nettleton, & Brummer, 2009) found that nonadditive expression of transcript levels may contribute to heterosis for biomass yield in alfalfa. Zhao (Zhao, Chai, & Liu, 2007), Meng (Meng, Ni, Wu, & Sun, 2005), Uźarowska et al.(Uźarowska et al., 2007) received the similar results.

Considering effects of DNA methylation on gene expression(Finnegan et al., 2000; Jacobsen et al., 2000; Rangwala & Richards, 2004), Hyper-methylation can lead to the gene silence, and demethylation can lead to the over-expression (Neves, Heslop-Harrison, & Viegas, 1995; Sardana, O'Dell, & Flavell, 1993), there may be a correlation between the DNA methylation and heterosis.

A negative correlation was found between DNA methylation and the economic characters of shell length, shell height, soft body weight and adductor muscle weight, Wan(Wan, 2008) found that, in the Chongqing mountainous cattle and their hybrids, methylation affected heart girth and body weight highly significantly, methylation and body slanting length significantly. Methylation content and heterosis rates of heart girth, body weight showed significant correlation, it just indicated indirectly that DNA methylation have a correlation with the gene expression. Zhang, Shiu, Cal, and Borevitz (2008) found that cytosine methylation alterations immolL-1ediately upstream or downstream of the gene were inversely correlated with the degree of expression variation for that gene. Jin found a direct relationship between cytosine methylation alteration and gene expression variation. So we can say DNA cytosine methylation alteration lead to the expression of heterosis in some extent.

The methylation of the *Scapharca broughtonii* from Korean and China are different, the cytosine methylation ratio of the filial generation will between the parents in theory, but the result was that the filial generation cytosine methylation ratio was lower than any parents. And four classes of patterns of cytosine methylation characterized by differences in degree of methylation between the hybrids and parental lines: (1) the same level of methylation in both parental lines and the hybrids, (2) the same level of methylation in either parent or hybrid,(3) an increased level of methylation in the hybrids compared to the parents, (4) a decreased level of methylation in the hybrids, and the number of demethylation sites were more than the number of the hypermethylation sites. Hepburn found the same result, Tsaftaris(A. Tsaftaris et al., 1997) found that hybrids were less methylated than inbreeds. It was further proposed that differential DNA methylation patterns in hybrids may play an important role in materializing heterosis

It has been widely recognized that, in animals, the inheritance of the epigenetic state through mitotic rounds of cell division is relatively faithful, in development (embryogenesis and gametogenesis) the epigenetic state is reset, that is, erased and reestablished; and parental epigenetic state in plants is often stably inherited to sexual progenies (Cubas, Vincent, & Coen, 1999; Monk, Boubelik, & Lehnert, 1987; Riddle & Richards, 2002). Penterman thought that demethylation processes may be the result of a deficiency in enzymatic maintenance after DNA replication or an active enzymatic process involving plant glycosylases with specific functions in genomic imprinting and to protect genes from potentially deleterious methylation(Grewal & Elgin, 2002; Grewal & Klar, 1996; Penterman et al., 2007). Grewal and Klar(Grewal & Klar, 1996) showed that the epigenetic modification of a reporter gene placed in the mating-type region of Schizosaccharomyces pombe could be inherited through mitosis and meiosis. Furthermore, they showed that loci influencing this process were, either directly or indirectly, involved with the organization of heterochromatin(Grewal & Klar, 1996). More recent work has shown that these modifiers include histone deacetylases, histone methyltransferases and other structural proteins associated with telomeres and centromeres (Grewal & Elgin, 2002).

Anyway, the change of the DNA cytosine methylation took place, though the mechanism is still unknown. Two different alleles brought together in the hybrid may create a combined allelic expression pattern in the hybrids. Alternatively, at some loci, allelic interaction or a change in the spectrum of trans-acting factors causes gene expression in the hybrid to deviate from simple additive allelic expression patterns of the parents(Birchler et al., 2003; Gibson & Weir, 2005).The patterns of the F1 DNA methylation experienced change and adjustments in order to coordinate the expression of the gene from the parents, most of the cytosine sites' methylation patterns could be inherited to next generation stably, some sites' methylation patterns experience hypermethylated and demethylated, the demethylation cause some gene expressing largely, the hypermethylation restrain the expression of some gene, with the participation of the environment, a new methylation patterns was formed which added together the parents' characters and can adapted to the circumstances very well, so the heterosis is the result of the differential expression of some gene, the DNA methylation also have to do with the expression of heterosis.

Generally speaking, after investigating the difference of methylation between the parents and the offspring, we can find out the special methylation sites which may play an important part in the expression of heterosis, and then using the methods of sequences analysis and genetic analysis to find out the inherent factor of heterosis. The genomes of the animals and plants contain a great quantity of CG dinucleotides, and the methylation sequences which control the expression of some functional gene also contain a great quantity of CG dinucleotides, so it is easy to find out the change regularity of the methylation.The advantages of MSAP include simplicity, rapidness, low cost, higher sensitivity, and high polymorphism, however, this technique also has some limitations associated with resolving power.

In some organisms, methylation also occurred in CAGs and CTGs. Furthermore, the types of non-methylation and inner-methylation of a single strand can not be distinguished as both HpaII and MspI are capable of recognizing the sites of non-methylation and inner-methylation of a single strand, thus revealing similar patterns of methylation following PCR amplification. Moreover, outer methylation of double strands can not be detected through MSAP analysis. For these reasons, actual levels of the cytosines methylation are likely to be significantly higher than those detected in this study. Notwithstanding some limits, the results here reported show this technique to be highly efficient for large-scale detection of cytosine methylation in *Scapharca broughtonii*. The ability to isolate and amplify these MSAP fragments may thus make possible direct identification of sequences which play a role during the expression of heterosis.

References

Birchler, J., Auger, D., & Riddle, N. (2003). In search of the molecular basis of heterosis. *The Plant Cell Online, 15*(10), 2236.

Cervera, M., Ruiz-Garcia, L., & Martinez-Zapater, J. (2002). Analysis of DNA methylation in Arabidopsis thaliana based on methylation-sensitive AFLP markers. *Molecular Genetics and Genomics, 268*(4), 543-552.

Chen, X., Ma, Y., Chen, F., Song, W., & Zhang, L. (2009). Analysis of DNA methylation patterns of PLBs derived from Cymbidium hybridium based on MSAP. *Plant Cell, Tissue and Organ Culture, 98*(1), 67-77.

Crow, J. (2000). The rise and fall of overdominance. *Plant breeding reviews, 17*, 225-258.

Cubas, P., Vincent, C., & Coen, E. (1999). An epigenetic mutation responsible for natural variation in floral symmetry. *Nature, 401*(6749), 157-161.

Finnegan, E., Peacock, W., & Dennis, E. (2000). DNA methylation, a key regulator of plant development and other processes. *Current opinion in genetics & development, 10*(2), 217-223.

Gibson, G., & Weir, B. (2005). The quantitative genetics of transcription. *TRENDS in Genetics, 21*(11), 616-623.

Grewal, S., & Elgin, S. (2002). Heterochromatin: new possibilities for the inheritance of structure. *Current opinion in genetics & development, 12*(2), 178-187.

Grewal, S., & Klar, A. (1996). Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis. *Cell, 86*(1), 95-102.

Horváth, E., Szalai, G., Janda, T., Páldi, E., Rácz, I., & Lásztity, D. (2002). Effect of vernalisation and azacytidine on the DNA methylation level in wheat (Triticum aestivum L. cv. Mv 15). *Acta Biologica Szegediensis, 46*(3-4), 35-36.

Jacobsen, S., Sakai, H., Finnegan, E., Cao, X., & Meyerowitz, E. (2000). Ectopic hypermethylation of flower-specific genes in Arabidopsis. *Current Biology, 10*(4), 179-186.

Lamkey, K., & Edwards, J. (1997). *The quantitative genetics of heterosis*.

Leonardi, A., Damerval, C., Hebert, Y., Gallais, A., & Vienne, D. (1991). Association of protein amount polymorphism (PAP) among maize lines with performances of their hybrids. *TAG Theoretical and Applied Genetics, 82*(5), 552-560.

Li, X., Wei, Y., Nettleton, D., & Brummer, E. (2009). Comparative gene expression profiles between heterotic and non-heterotic hybrids of tetraploid Medicago sativa. *BMC Plant Biology, 9*(1), 107.

Li, Z., Luo, L., Mei, H., Wang, D., Shu, Q., Tabien, R., ... Khush, G. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics, 158*(4), 1737.

Lippman, Z., & Zamir, D. (2007). Heterosis: revisiting the magic. *TRENDS in Genetics, 23*(2), 60-66.

Lu, G., Wu, X., Chen, B., Gao, G., Xu, K., & Li, X. (2006). Detection of DNA methylation changes during seed germination in rapeseed (Brassica napus). *Chinese Science Bulletin, 51*(2), 182-190.

Luff, B., Pawlowski, L., & Bender, J. (1999). An inverted repeat triggers cytosine methylation of identical sequences in Arabidopsis. *Molecular Cell, 3*(4), 505-511.

Luo, L., Li, Z., Mei, H., Shu, Q., Tabien, R., Zhong, D., ... Paterson, A. (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. *Genetics, 158*(4), 1755.

Meng, F., Ni, Z., Wu, L., & Sun, Q. (2005). Differential gene expression between cross-fertilized and self-fertilized kernels during the early stages of seed development in maize. *Plant science, 168*(1), 23-28.

Monk, M., Boubelik, M., & Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development, 99*(3), 371.

Neves, N., Heslop-Harrison, J., & Viegas, W. (1995). rRNA gene activity and control of expression mediated by methylation and imprinting during embryo development in wheat x rye hybrids. *TAG Theoretical and Applied Genetics, 91*(3), 529-533.

Paszkowski, J., & Whitham, S. (2001). Gene silencing and DNA methylation processes. *Current opinion in plant biology, 4*(2), 123-129.

Penterman, J., Zilberman, D., Huh, J., Ballinger, T., Henikoff, S., & Fischer, R. (2007). DNA demethylation in the Arabidopsis genome. *Proceedings of the National Academy of Sciences, 104*(16), 6752.

Rangwala, S., & Richards, E. (2004). The value-added genome: building and maintaining genomic cytosine methylation landscapes. *Current opinion in genetics & development, 14*(6), 686-691.

Reif, J., Warburton, M., Xia, X., Hoisington, D., Crossa, J., Taba, S., ... Melchinger, A. (2006). Grouping of accessions of Mexican races of maize revisited with SSR markers. *TAG Theoretical and Applied Genetics, 113*(2), 177-185.

Reyna-Lopez, G., Simpson, J., & Ruiz-Herrera, J. (1997). Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Molecular and General Genetics MGG, 253*(6), 703-710.

Riddle, N., & Richards, E. (2002). The control of natural variation in cytosine methylation in Arabidopsis. *Genetics, 162*(1), 355.

Romagnoli, S., Maddaloni, M., Livini, C., & Motto, M. (1990). Relationship between gene expression and hybrid vigor in primary root tips of young maize (Zea mays L.) plantlets. *TAG Theoretical and Applied Genetics, 80*(6), 769-775.

Sardana, R., O'Dell, M., & Flavell, R. (1993). Correlation between the size of the intergenic regulatory region, the status of cytosine methylation of rRNA genes and nucleolar expression in wheat. *Molecular and General Genetics MGG, 236*(2), 155-162.

Soppe, W., Jacobsen, S., Alonso-Blanco, C., Jackson, J., Kakutani, T., Koornneef, M., & Peeters, A. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Molecular Cell, 6*(4), 791-802.

Sun, Q., Wu, L., Ni, Z., Meng, F., Wang, Z., & Lin, Z. (2004). Differential gene expression patterns in leaves between hybrids and their parental inbreds are correlated with heterosis in a wheat diallel cross. *Plant science, 166*(3), 651-657.

Tariq, M., & Paszkowski, J. (2004). DNA and histone methylation in plants. *TRENDS in Genetics, 20*(6), 244-251.

Tsaftaris, A., Kafka, M., Polidoros, A., & Tani, E. (1997). *Epigenetic changes in maize DNA and heterosis*.

Tsaftaris, S. (2006). Molecular aspects of heterosis in plants. *Physiologia plantarum, 94*(2), 362-370.

Uźarowska, A., Keller, B., Piepho, H., Schwarz, G., Ingvardsen, C., Wenzel, G., & Lübberstedt, T. (2007). Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height. *Plant Molecular Biology, 63*(1), 21-34.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T., Hornes, M., ... Kuiper, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research, 23*(21), 4407.

Wan, Y. (2008). Study on the Relationship beween DNA Methylation and Heterosis in Beef Cattle. *Master paper, Southwest University,Chongqing,China(in Chinese)*, 25-33.

Xiao, J., Li, J., Yuan, L., & Tanksley, S. (1995). Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. *Genetics, 140*(2), 745.

Xiong, L., Xu, C., Saghai Maroof, M., & Zhang, Q. (1999). Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Molecular and General Genetics MGG, 261*(3), 439-446.

Xu, M., Li, X., & Korban, S. (2000). AFLP-based detection of DNA methylation. *Plant Molecular Biology Reporter, 18*(4), 361-368.

Yu, S., Li, J., Xu, C., Tan, Y., Gao, Y., Li, X., ... Maroof, M. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences of the United States of America, 94*(17), 9226.

Zhang, X., Shiu, S., Cal, A., & Borevitz, J. (2008). Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays. *PLoS Genetics, 4*(3).

Zhao, X., Chai, Y., & Liu, B. (2007). Epigenetic inheritance and variation of DNA methylation level and pattern in maize intra-specific hybrids. *Plant science, 172*(5), 930-938.

Zilberman, D., & Henikoff, S. (2007). Genome-wide analysis of DNA methylation patterns. *Development, 134*(22), 3959.

Journal of Molecular Biology Research; Vol. 5, No. 1; 2015
ISSN 1925-430X E-ISSN 1925-4318
Published by Canadian Center of Science and Education

Reviewer Acknowledgements

Journal of Molecular Biology Research wishes to acknowledge the following individuals for their assistance with peer review of manuscripts for this issue. Their help and contributions in maintaining the quality of the journal is greatly appreciated.

Journal of Molecular Biology Research is recruiting reviewers for the journal. If you are interested in becoming a reviewer, we welcome you to join us. Please find the application form and details at http://www.ccsenet.org/reviewer and e-mail the completed application form to jmbr@ccsenet.org.

Reviewers for Volume 5, Number 1

Chandrasekhar Natarajan, University of Nebraska-Lincoln, School of Biological sciences, Lincoln, USA

Deovrat Begde, Dr. Ambedkar College, India

Ding Yun, Janelia Farm Research Campus, HHMI, USA

Fernando Cardona, Genomics and Proteomics Department. Institute of Biomedicine of Valencia-CSIC, Spain

Guoku Hu, Creighton University, USA

Haoyu Si, University of Maryland Baltimore County, USA

Jianing Xu, University of Georgia, USA

Jiannan Guo, HHMI/University of Pennsylvania, USA

José Luis Fernández, Veterinary Faculty of Extremadura University (UNEX), Cáceres, Sapin

Junjie Xu, University of Texas Southwestern Medical Center, USA

Na Luo, Indiana University, USA

Nimrat Chatterjee, Baylor College of Medicine, USA

Prasad Siddavatam, Thermo Fisher Scientific, USA

Sandeep Pawar, University of Chicago, USA

Taraka Donti, Baylor College of Medicine, USA

Yibin Lin, University of Texas Medical School at Houston, USA

Call for Manuscripts

Journal of Molecular Biology Research is an international, double-blind peer-reviewed, open-access journal, published by the Canadian Center of Science and Education. It publishes original research, applied, and educational articles in all areas of molecular biology. The journal is available in electronic form in conjunction with its print edition. All articles and issues are available for free download online.

We are seeking submissions for forthcoming issues. All manuscripts should be written in English. Manuscripts from 3000-8000 words in length are preferred. All manuscripts should be prepared in MS-Word format, and submitted online, or sent to: jmbr@ccsenet.org

Paper Selection and Publishing Process

a) Upon receipt of a submission, the editor sends an e-mail of confirmation to the submission's author within one to three working days. If you fail to receive this confirmation, your submission e-mail may have been missed.

b) Peer review. We use a double-blind system for peer review; both reviewers' and authors' identities remain anonymous. The paper will be reviewed by at least two experts: one editorial staff member and at least one external reviewer. The review process may take two to three weeks.

c) Notification of the result of review by e-mail.

d) If the submission is accepted, the authors revise paper and pay the publication fee.

e) After publication, the corresponding author will receive two hard copies of the journal, free of charge. If you want to keep more copies, please contact the editor before making an order.

f) A PDF version of the journal is available for download on the journal's website, free of charge.

Requirements and Copyrights

Submission of an article implies that the work described has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the authorities responsible where the work was carried out, and that, if accepted, the article will not be published elsewhere in the same form, in English or in any other language, without the written consent of the publisher. The editors reserve the right to edit or otherwise alter all contributions, but authors will receive proofs for approval before publication.

Copyrights for articles are retained by the authors, with first publication rights granted to the journal. The journal/publisher is not responsible for subsequent uses of the work. It is the author's responsibility to bring an infringement action if so desired by the author.

More Information

E-mail: jmbr@ccsenet.org

Website: www.ccsenet.org/jmbr

Paper Submission Guide: www.ccsenet.org/submission

Recruitment for Reviewers: www.ccsenet.org/reviewer

The journal is peer-reviewed
The journal is open-access to the full text
The journal is included in:

EBSCOhost
Google Scholar
LOCKSS
NewJour (Georgetown University Library)
PKP Open Archives Harvester
ProQuest
SHERPA/RoMEO
Standard Periodical Directory
Ulrich's

Journal of Molecular Biology Research

Annual

Publisher Canadian Center of Science and Education
Address 1120 Finch Avenue West, Suite 701-309, Toronto, ON., M3J 3H7, Canada
Telephone 1-416-642-2606
Fax 1-416-642-2608
E-mail jmbr@ccsenet.org
Website www.ccsenet.org/jmbr

ISSN 1925-430X