

‘COGNITIVE GENES’ REVEAL HIGHER CODON COMPLEXITY THAN ‘SOMATIC GENES’

Christoph S. Herrmann^{1*} and *Wolfgang S. Herrmann*²

¹Otto-von-Guericke-University, Dept. for Biological Psychology,
Magdeburg, Germany

²Johann-Wolfgang-Goethe-University, Dept. for Modern Languages,
Frankfurt, Germany

Abstract

In this article we want to apply the concept of complexity to the analysis and comparison of genes. A multitude of genes has been identified coding somatic function. Recently the analysis of mental disorders yielded insights about genes coding cognitive functions. According to the theory of evolution they evolved from other genes through mutation. Therefore, ‘cognitive genes’ and ‘somatic genes’ should differ in their coding reflecting these mutations. We investigated ‘cognitive genes’ and ‘somatic genes’ and demonstrated that their codon usage differs significantly. ‘Somatic genes’ are coded in accordance with the average codon usage of a species — ‘cognitive genes’ differ from it, i.e. they reveal a higher codon complexity. This increased complexity might reflect the mutations which occurred during evolution.

Keywords: human, mouse, species specific codon usage

1. Introduction

The human body with its interaction of myriad subsystems is a perfect example for a complex dynamical system. Especially the brain as one such subsystem has been viewed from the dynamical systems’ point of view (Başar, 1990; Haken, 2002). Applying tools from dynamical systems theory has offered new insights for the understanding of healthy as well as pathological electroencephalograms (beim Graben et al., 2000; Quiroga et al., 2000; Freeman, 2003). But even the DNA (desoxy ribonucleic acid) which codes this whole dynamical system exhibits features of dynamical systems (Vinogradov, 2004). Recently, terms like ‘organismal complexity’, ‘gene complexity’, and ‘protein complexity’ have been coined (Yang et al., 2003; He and Zhang, 2005) and tools like entropy measures have been applied to gene analysis (Furlanello et al., 2003; Eng et al., 2004; Schug et al., 2005). A whole issue of the journal ‘Chaos’ has been dedicated to the application of dynamical systems’ knowledge to molecular and cell biology (Tyson and Mackey, 2001). Here, we want

*E-mail address: christoph.herrmann@nat.uni-magdeburg.de, <http://www.uni-magdeburg.de/cherrman>
(Corresponding author)

to introduce a new perspective onto genes that code cognitive functions from the view-point of dynamical systems, namely complexity.

Since the pioneering work of Mendel a wealth of experimental data have revealed how living creatures inherit their physical features from their parents (Mc Kusick, 1992). Meanwhile even the inheritance of clinical syndromes has been demonstrated and in some patients genes could be identified which if mutated result in mental disorders such as Alzheimer and Parkinson (Selkoe, 1996; Duvoisin, 1986). Recently, even the analysis of speech disorders revealed a gene which is mutated in patients with a severe language disorder, called FoxP2 (Fisher et al., 1998; Lai et al., 2001). The gene is considered the first language gene (Pinker, 2001). The theory of evolution implies that those genes which code higher cognitive functions must have evolved from predecessor genes through mutation. This raises the question of what makes a gene a ‘cognitive gene’¹. Can one actually find more mutations in ‘cognitive genes’ than in ‘somatic genes’?

One possible strategy to address this issue is to compare human genes with genes of our evolutionary predecessors. For example, the language gene FoxP2 is also found in chimpanzees and mice (Enard et al., 2002a,b). A close comparison of the FoxP2 genes revealed that only 6 of 715 amino acids differ between humans and mice. This seems counterintuitive given the obviously different ability to use language. If, however, the differences in codon usage are examined, a total of 13 differences was found between humans and chimpanzees and 115 differences between humans and mice (cf. Fig. 1). Most of these differences in codon usage did not result in a change of amino acids due to the redundant coding of genes and are considered synonymous. However, it has been argued recently that even synonymous changes in codon usage can be meaningful for the coded function (Duan et al., 2003; Xia, 1996; Richard and Beckmann, 1995). Thus, analyzing the codon usage of ‘cognitive genes’ and ‘somatic genes’ might yield important differences.

Another way of testing whether ‘cognitive genes’ reveal more mutations than ‘somatic genes’ is to investigate the average coding used for each amino acid. Each species has a species-specific codon usage for every amino acid. If ‘cognitive genes’ showed more mutations than ‘somatic genes’ then ‘cognitive genes’ should deviate from the species-specific average coding more than ‘somatic genes’. In order to coin a term for this potentially different codon behaviour, ‘cognitive genes’ should exhibit a higher *codon complexity* (see below for a definition) than ‘somatic genes’.

We selected 24 genes from the NCBI nucleotide database (www.ncbi.nih.gov) which are associated with cognitive diseases or functions, e.g. neurotransmitter (NT) receptors. These will be considered ‘cognitive genes’ (cf. Tab. 1). In addition, another 24 genes were selected which are associated with basic physical functions, like blood pressure control or metabolism. These will be considered ‘somatic genes’ (cf. Tab. 2).

¹Of course, a ‘cognitive gene’ like FoxP2 does not code for a cognitive function like speech but rather for a sequence of amino acids (a protein) which is crucial for speech or speech development. For this reason the term ‘cognitive gene’ will be used in apostrophes. Along the same lines, ‘somatic genes’ like tumor-causing oncogenes don’t code cancer but rather proteins involved in controlling cell division and differentiation. Thus, also ‘somatic genes’ will be used in apostrophes.

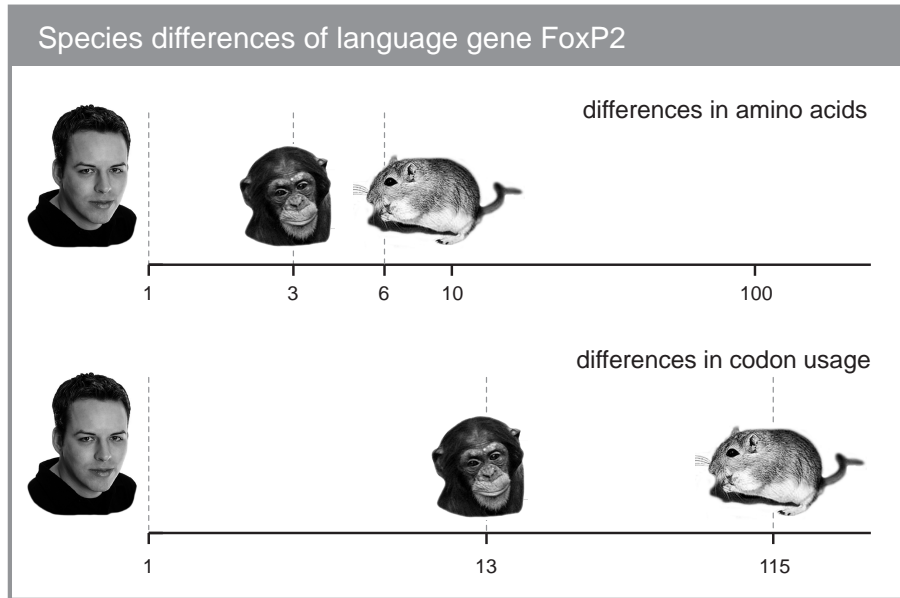


Figure 1. Differences of the language gene FoxP2 between humans, chimpanzees, and mice. Chimpanzees and mice differ from humans in only 3 and 6 amino acids, respectively. However, when comparing codon usage, 13 and 115 differences can be found, respectively.

2. Methods

2.1. Comparing Codons and Amino Acid Sequences

In order to demonstrate that codons better differentiate between species, we compared the number of different amino acids and the number of different codons for all 48 genes between humans (*homo sapiens*, HS) and mice (*mus musculus*, MM). Sequences were aligned using ClustalW (Thompson et al., 1994). If the alignments were non-overlapping at the 5' or 3' ends of the open reading frames, respectively, only the overlapping sequences which ClustalW found to be homologous were further analyzed and the overlap was truncated. An ANOVA was computed on the number of differences to reveal whether the method of comparison (acid or codon usage) yielded significant differences.

2.2. Comparing ‘Cognitive Genes’ and ‘Somatic Genes’

In order to compare the codon usage for ‘cognitive genes’ and ‘somatic genes’ in humans and mice, we needed an average value for the codon usage in both species. Such values have been compiled in tables of species-specific codon usage (SSCU) and are available on the internet (www.kazusa.or.jp) (Nakamura et al., 1999). For each amino acid it is stated which of the alternative codons is used how frequently, i.e. alanine is coded by GCA in 22.5%, by GCC in 40.4%, by GCG in 10.7%, and by GCT in 26.3% of all cases (cf. complete list for humans in Fig. 2). In a next step we had to compare the codon usage of an individual gene to this species specific codon usage. We computed Pearson’s correlations of the SSCU with

Table 1. The 24 ‘cognitive genes’ used and their accession numbers in the NCBI nucleotide database.

Abbrev.:	Description:	Acc. # HS	Acc. # MM	Function/disorder:
ADSL	adenylsuccinate lyase	AF067854	XM122934	autism
APP	amyloid β precursor protein	NM000484	XM128362	Alzheimer’s disease
CHRM1	cholinergic receptor, muscarinic 1	XM170669	NM007698	NT receptor
CHRNA3	cholin. rec., nicotinic α polypeptide 3	NM000743	NM145129	NT receptor
CREB1	cAMP respons. elem. bind. protein 1	XM028941	M95106	long-term memory
CREB2	cAMP resp. elem. bind. protein 2	XM017578	XM139474	long-term memory
DAB1	disabled homolog 1	NM021080	NM010014	brain development
DARPP32	dopamine/cAMP-regul. phosphoprot.	AF464196	XM126562	Parkinson’s disease
FOXP2	forkhead helix transcription factor	AY144615	NM053242	language
GABA	GABA-A receptor δ	BC033801	BC031762	NT receptor
GPHN	gephyrin	NM020806	XM126999	NT receptor-clustering
IMMP2	inner mitochondr. membr. peptidase 2	NM032549	AF359564	Tourette syndrome
MAPT	microtubule-associated protein tau	XM038684	NM010838	Parkinson’s disease
MAP2K1	mitogen-activated protein kinase 1	NM002755	NM008927	long-term memory
MAP2K2	mitogen-activated protein kinase 2	NM030662	NM023138	long-term memory
MECP1	methyl CpG binding protein 1	NM020699	NM139304	Rett syndrome
MECP2	methyl CpG binding protein 2	XM010162	NM010788	Rett syndrome
NDN	neddin homolog	NM002487	BC015268	Prader-Willi syndrome
NPC1	Niemann-Pick disease type C1	NM000271	NM008720	dementia
PSN1	presenilin 1	AJ008005	BC030409	Alzheimer’s disease
PSN2	presenilin 2	XM002127	NM011183	Alzheimer’s disease
RP42	telencephalic RP42	NM020640	AF198092	autism susceptibility
SLC6A6	solute carrier family 6	AF346763	NM009320	NT transporter
PRNP	propion prion	NM000311	NM011170	encephalopathy

a gene’s CU. These correlations were computed across the 59 codons which are encoded by more than one alternative coding sequence, i.e. methionine, tryptophan, and the three stop codons were excluded. The resulting r^2 of the correlation is 1 when the gene uses the identical codon frequencies as the species average and it is lower the more the actual coding of a gene deviates from the species average coding. Since low predictability is one of the many ways to compute complexity (Schreiber, 1999), we will consider $(1 - r^2)$ as a gene’s codon complexity because a gene’s codon usage becomes less predictable with decreasing r^2 . I.e., if a gene is coded exactly as the mean (high r^2), its codon complexity will be low. If, however, its codon usage differs significantly from the species-specific average (low r^2) its codon complexity will be high. ANOVAs were computed on these correlations for humans and mice to test whether the type of gene (cognitive vs. somatic) yielded any significant differences.

3. Results

3.1. Comparing Codons and Amino Acid Sequences

An ANOVA comparing all 48 genes yielded a significant difference for the method used to compare the genes between humans and mice ($F(1,47)=63.27, p<0.0001$). Comparisons of amino acids revealed 14.87% differences between the 48 genes for humans and mice while comparing codon usage revealed 36.61% differences.

3.2. Comparing ‘Cognitive Genes’ and ‘Somatic Genes’

Fig. 2 shows the species-specific and gene-specific codon usage (GSCU) of FoxP2 in humans and mice for all informative amino acids. A small difference between these two measures reflects a low codon complexity, i.e. the gene is coded in line with the average codon usage of the species. A large difference, however, indicates a high codon complexity.

The ANOVA of the correlations across all human genes yielded a significant main effect for the type of gene ($F(1,23)=5.3$, $p=0.026$). ‘Somatic genes’ revealed a lower codon complexity ($r^2=.67$) as compared to ‘cognitive genes’ ($r^2=.53$, cf. Fig. 3). The same ANOVA of the correlations across all mouse genes also yielded a significant main effect for the type of gene ($F(1,23)=9.53$, $p=0.003$). ‘Somatic genes’ revealed a lower codon complexity ($r^2=.74$) as compared to ‘cognitive genes’ ($r^2=.57$). Tab. 3 illustrates all correlations across the 24 ‘cognitive genes’ and the 24 ‘somatic genes’.

Code	HS			MM			Code	HS			MM		
	SS	GS	DI	SS	GS	DI		SS	GS	DI	SS	GS	DI
Agca	22,5	44,2	21,7	22,6	33,3	10,7	Pcca	27,4	36,4	9,0	28,2	29,5	1,3
Agcc	40,5	25,6	14,9	38,5	38,1	0,4	Pccc	32,8	15,9	16,9	30,7	27,3	3,4
Agcg	10,7	4,7	6,1	10,1	7,1	2,9	Pccg	11,5	6,8	4,7	10,8	6,8	4,0
Agct	26,3	25,6	0,7	28,8	21,4	7,4	Pcct	28,4	40,9	12,5	30,3	36,4	6,0
Ctgc	55,2	42,9	12,3	53,7	28,6	25,1	Qcaa	25,4	33,0	7,7	25,2	28,6	3,4
Ctgt	44,8	57,1	12,3	46,3	71,4	25,1	Qcag	74,6	67,0	7,7	74,8	71,4	3,4
Dgac	53,9	54,5	0,6	55,9	66,7	10,7	Raga	20,1	26,9	6,8	20,5	15,4	5,2
Dgat	46,1	45,5	0,6	44,1	33,3	10,7	Ragg	20,1	19,2	0,9	21,1	30,8	9,7
Egaa	41,6	63,4	21,8	40,1	54,8	14,7	Rcga	11,3	30,8	19,5	12,4	26,9	14,6
Egag	58,4	36,6	21,8	59,9	45,2	14,7	Rcgc	19,2	3,8	15,4	18,0	7,7	10,3
Fttc	54,7	20,0	34,7	57,1	30,0	27,1	Rcgg	21,0	3,8	17,1	19,3	11,5	7,7
Fttt	45,3	80,0	34,7	42,9	70,0	27,1	Rcgt	8,3	15,4	7,1	8,7	7,7	1,0
Ggga	24,6	56,7	32,1	25,5	46,7	21,2	Sagc	24,3	22,4	1,9	24,5	25,0	0,5
Gggc	34,4	30,0	4,4	33,6	30,0	3,6	Sagt	15,0	23,7	8,7	15,1	21,1	6,0
Gggg	24,7	6,7	18,1	23,4	13,3	10,1	Stca	14,7	19,7	5,0	13,8	19,7	5,9
Gggt	16,3	6,7	9,6	17,5	10,0	7,5	Stcc	21,9	13,2	8,8	22,0	13,2	8,8
Hcac	58,9	50,0	8,9	60,4	37,5	22,9	Stcg	5,7	5,3	0,4	5,6	6,6	1,0
Hcat	41,1	50,0	8,9	39,6	62,5	22,9	Stct	18,4	15,8	2,6	19,0	14,5	4,6
Iata	16,0	37,0	21,0	15,1	33,3	18,3	Taca	27,9	30,0	2,1	28,8	34,1	5,4
Iatc	48,5	22,2	26,3	51,5	40,7	10,7	Tacc	36,2	30,0	6,2	35,8	29,3	6,5
Iatt	35,4	40,7	5,3	33,5	25,9	7,6	Tacg	11,7	2,5	9,2	11,1	7,3	3,8
Kaaa	42,1	60,0	17,9	38,1	48,0	9,9	Tact	24,2	37,5	13,3	24,4	29,3	4,9
Kaag	57,9	40,0	17,9	61,9	52,0	9,9	Vgta	11,4	21,6	10,2	11,4	13,2	1,8
Lcta	7,0	11,1	4,1	7,7	9,7	2,0	Vgtc	23,7	24,3	0,6	25,1	28,9	3,8
Lctc	19,6	13,9	5,7	20,0	18,1	2,0	Vgtg	47,2	40,5	6,6	47,0	50,0	3,0
Lctg	40,7	19,4	21,3	40,5	22,2	18,3	Vgtt	17,7	13,5	4,2	16,5	7,9	8,6
Lctt	12,8	22,2	9,4	12,7	16,7	4,0	Ytac	56,5	50,0	6,5	58,6	60,0	1,4
Ltta	7,3	20,8	13,6	6,1	16,7	10,6	Ytat	43,5	50,0	6,5	41,4	40,0	1,4
Lttg	12,6	12,5	0,1	12,9	16,7	3,7							
Naac	53,9	23,3	30,5	57,8	41,4	16,4							
Naat	46,1	76,7	30,5	42,2	58,6	16,4							

Figure 2. Codon usage for FoxP2 in humans (HS) and in mice (MM). The word in the first column codes the amino acid (capital letter for the name and lower case for the three bases). Species-specific codon usage in percent per amino acid are given in column SS. The gene-specific (GS) usage in FoxP2 and the difference (DI) between the two reveal that this ‘cognitive gene’ varies in coding between humans and mice. Gray boxes mark those amino acids in which GS differs from SS by more than 20%.

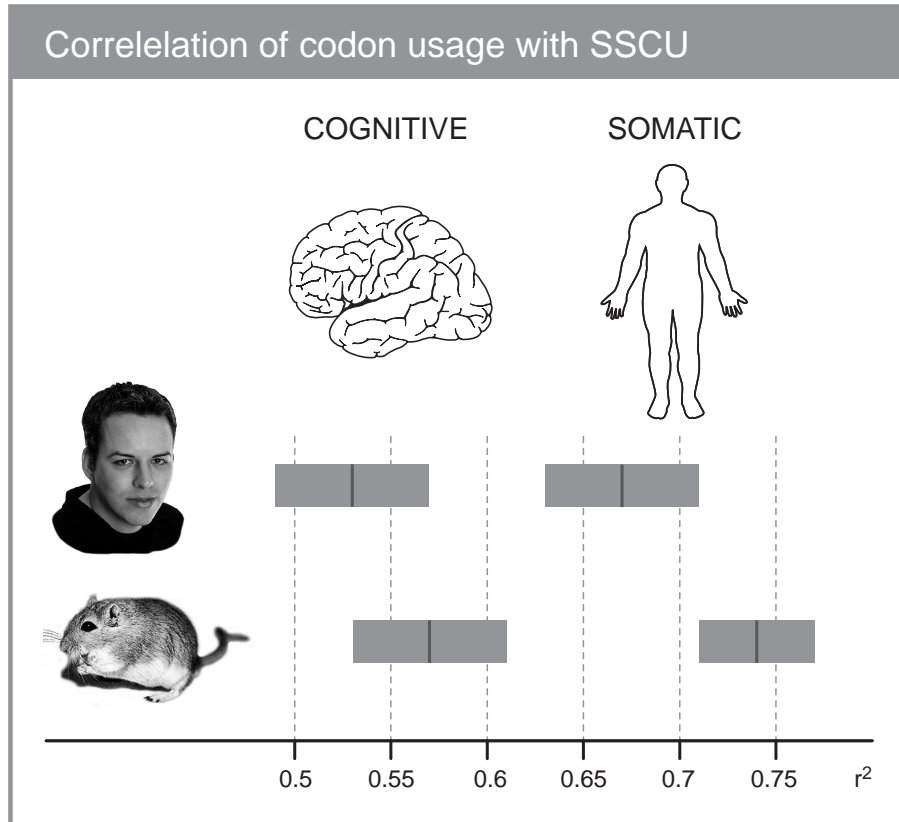


Figure 3. Correlation coefficients (r^2) of GSCU with SSCU in humans (HS) and in mice (MM). Gray bars represent standard error of the mean. In both species, ‘somatic genes’ are coded with higher correlation (lower codon complexity) than ‘cognitive genes’. This indicates that codon complexity might be a meaningful parameter.

4. Discussion

The main result of our study is that the codon usage of ‘cognitive genes’ reveals a significantly higher codon complexity as compared to ‘somatic genes’. This is true for both human and mouse genes. In addition, we were able to replicate the known fact that comparing codon usage yields more differences between species than comparing amino acids. Since our findings demonstrate that the codon complexity of genes differs significantly between ‘cognitive genes’ and ‘somatic genes’, it seems necessary to compare codon usage instead of amino acids when comparing genes between species. Obviously, this yields more differences between species and our results support the notion that these differences are not as redundant as mostly assumed until today. On the contrary, codon usage seems an important factor for protein synthesis (Duan et al., 2003) and probably results from selection in order to minimize the effects of errors, e.g. mutations (Freeland et al., 2000).

Our findings raise the question why ‘cognitive genes’ and ‘somatic genes’ might be

coded differently. One potential explanation is offered by the following argumentation. Genes which are expressed broadly follow SSCU more closely than those which are expressed tissue-specific (Urrutia and Hurst, 2001). For ‘cognitive genes’ with their high codon complexity this would mean that they are expressed less frequently than ‘somatic genes’. Under the assumption that many ‘cognitive genes’ are expressed mainly in neural tissue while most ‘somatic genes’ are expressed throughout the body, this implication of our findings seems plausible. In addition, it has been argued that evolutionary changes occur more frequently in genes which are expressed later during ontogeny (Castillo-Davis and Hartl, 2002). Under the assumption that genes showing a low codon complexity are those in which only few evolutionary changes appeared, it follows that ‘cognitive genes’ are those that underwent more evolutionary changes. This goes well in line with such genes being expressed later in ontogeny, since the somatic development of organs precedes the subsequent cognitive developments.

An alternative explanation might be found in antisense transcription. Recently, short antisense sequences were found to be transcribed from the antisense strand of human DNA which was formerly believed not to be transcribed (Yelin et al., 2003). This phenomenon was initially found in prokaryotes (species without cell nuclei) but not in humans (Wagner and Simons, 1994). However, meanwhile antisense transcripts have been discovered also in many eukaryotes (species with cell nuclei) including humans (Kumar and Carmichael, 1998; Chen et al., 2004). Since the mutations of many ‘cognitive genes’ such as FoxP2 show less changes in transcribed amino acids than in codons during evolution, it seems plausible to assume that potentially present antisense transcripts undergo relevant changes. Such antisense transcripts are believed to influence gene expression (Knee and Murphy, 1997) at transcriptional as well as post-transcriptional level (Lipman, 1997). Thus, one might assume that especially the *expression* of ‘cognitive genes’ is affected by their codon usage.

Still another potential explanation for the difference between ‘cognitive genes’ and ‘somatic genes’ can be found in their transcription efficiency. Genes that follow SSCU strictly were found to transcribe more efficiently than those which follow SSCU only loosely (Sharp et al., 1995). This might indicate that ‘cognitive genes’, which exhibit higher codon complexity in their codon usage than ‘somatic genes’, are transcribed less efficiently than somatic genes. The different codons used for the same amino acids are believed to code different secondary structures of the proteins (Oresic and Shalloway, 1998). It remains to be seen how these structural differences result in the observed functional differences in case of the ‘cognitive genes’.

Comparing codon complexity may prove useful in many future applications. Our findings might be of diagnostic relevance when evaluating the codon complexity of ‘somatic genes’ which indicate diseases, e.g. in the case of the p53 gene responsible for cancer (Vogelstein and Kinzler, 1994). It seems plausible to assume that a change in codon complexity may occur before the actual onset of the disease resulting in the possibility of early intervention.

Table 2. The 24 ‘somatic genes’ used and their accession numbers in the NCBI nucleotide database.

Abbrev.:	Description:	Acc. # HS	Acc. # MM	Function:
AGTRL1	angiotensin receptor-like 1	NM005161	NM011784	blood pressure
AR	androgen receptor	NM000044	NM013476	sex function
BDKRB1	bradykinin receptor B1	NM000710	NM007539	blood pressure
EGF	epidermal growth factor	AF442487	AY032924	skin growth
ESR2	estrogen receptor 2	NM001437	NM010157	sex function
FANCC	Fanconi anemia, complement. group C	XM047190	NM007985	Fanconi anemia
GBA2	glucosidase, bile acid 2	NM020944	NM172692	muscle atrophy
GCG	glucagon	XM029322	AK007911	blood sugar
GHR	growth hormone receptor	XM003896	NM010284	body growth
KCNJ6	potassium channel 6	XM048829	NM010606	cell function
IGF1	insulin-like growth factor	NM000618	NM010512	body growth
MATP	membrane assoc. transporter protein	NM016180	NM053077	pigmentation
MCRS1	microspherule protein 1	NM006337	NM016766	immune function
NOS1	nitric oxide synthase 1	NM000620	NM008712	vasoconstriction
PAH	phenylalanine hydroxylase	NM000277	NM008777	metabolism
PTPN6	tyrosine phosphatase non-receptor 6	NM080549	NM013545	muscle coordination
P53	protein 53	AF520698	AF520699	Cancer protection
SPC	xeroderma pigmentosum, group C	NM004628	NM009531	UV protection
TAP1	peptide transporter	L21208	NM013683	Morbus Bechterew
TAP2	ABC transporter	AB073779	NM011530	Morbus Bechterew
TCF4	transcription factor 4	NM003199	NM013685	body growth
TCOF1	Treacher Collins-Franceschetti 1	NM000356	NM011552	cleft lip palate
TITF1	thyroid transcription factor 1	NM003317	NM009385	thyroid function
TSHR	TSH receptor	NM000369	NM011648	thyroid stimulation

5. Conclusions

On the one hand, mutations of genes have led to cognition during the course of evolution and the more recent genes code functions which are more adaptive to our environment and might be considered ‘more intelligent’. On the other hand, the same mutations led not only to more complex and more flexible systems, but, as has been demonstrated above, also to genes with higher codon complexity. This deviation from the species-specific codon usage may introduce not only adaptation to our environment but also make the genes more vulnerable for pathological mutations. Therefore, cognitive functions and cognitive disorders might just be two sides of the same coin: recent mutations of genes exhibiting relatively low codon complexity into ‘cognitive genes’ with high codon complexity.

Acknowledgements

We express our thanks to Prof. Dr. Ulrich Brandt and Dr. Stefan Kerscher (Dept. Bioenergetics, Frankfurt University), Prof. Dr. Jürgen Winkler (Dept. of Neuroanatomy, Frankfurt University), and Chi-Vinh Duong (Max-Planck-Institute for Brain Research, Frankfurt) for valuable discussions. We thank Andrea Gast-Sandmann (Max-Planck-Institute of Cognitive Neuroscience, Leipzig) for the illustrations and Sylvio Tüpke (Max-Planck-Institute for Evolutionary Anthropology, Leipzig) for the picture of the chimpanzee.

Table 3. Correlation (r^2) of SSCU with GSCU for the investigated 'cognitive genes' (left column) and 'somatic genes' (right column) in humans (HS) and mice (MM). On average, 'somatic genes' reveal a higher correlation (i.e. lower codon complexity) than 'cognitive genes', both in humans (0.66 vs. 0.53) and in mice (0.73 vs. 0.57).

Cog. genes:	HS:	MM:	Som. genes:	HS:	MM:
GPHN	0.08	0.10	BDKRB1	0.10	0.67
CREB2	0.12	0.05	KCNJ6	0.15	0.71
RP42	0.15	0.17	TITF1	0.42	0.17
CREB1	0.30	0.37	GHR	0.48	0.45
IMMP2	0.38	0.36	IGF1	0.55	0.59
DAB1	0.42	0.33	P53	0.57	0.84
FOXP2	0.44	0.61	AGTRL1	0.59	0.71
MAPT	0.52	0.61	FANCC	0.64	0.71
GABA	0.53	0.66	TCF4	0.64	0.81
NDN	0.53	0.67	PAH	0.67	0.76
DARPP32	0.55	0.55	TCOF1	0.69	0.76
SLC6A6	0.56	0.71	MCRS1	0.69	0.71
PRNP	0.59	0.55	GBA2	0.71	0.77
MECP1	0.61	0.61	GCG	0.76	0.58
MECP2	0.62	0.67	TAP2	0.77	0.79
CHRM1	0.62	0.67	MATP	0.81	0.79
PSN2	0.62	0.74	NOS1	0.81	0.85
PSN1	0.64	0.72	TAP1	0.81	0.76
MAP2K1	0.66	0.71	EGF	0.85	0.87
CHRNA3	0.71	0.62	AR	0.87	0.90
ADSL	0.72	0.76	ESR2	0.88	0.88
MAP2K2	0.72	0.64	TSHR	0.88	0.81
APP	0.81	0.79	SPC	0.88	0.88
NPC1	0.85	0.85	PTPN6	0.90	0.94
Average:	0.53	0.57		0.66	0.73

References

- Başar, E. (Ed.), 1990. *Chaos in Brain Function*. Springer, Berlin.
- beim Graben, P., Saddy, J. D., Schlesewsky, M., Kurths, J., 2000. Symbolic dynamics of event-related brain potentials. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **62** (4 Pt B), 5518–41.
- Castillo-Davis, C. I., Hartl, D. L., 2002. Genome evolution and developmental constraint in *Caenorhabditis elegans*. *Mol Biol Evol* **19** (5), 728–735.
- Chen, J., Sun, M., Kent, W. J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R. Z., Rowley, J. D., 2004. Over 20% of human transcripts might form sense-antisense pairs. *Nucleic Acids Res* **32** (16), 4812–20.
- Duan, J., Wainwright, M. S., Comeron, J. M., Saitou, N., Sanders, A. R., Gelernter, J., Gej-

- man, P. V., 2003. Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Human Molecular Genetics* **12** (3), 205–216.
- Duvoisin, R. C., 1986. Genetics of parkinson's disease. *Adv Neurol* **45**, 307–312.
- Enard, W., Khaitovich, P., Klose, J., Zöllner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G. M., Bontrop, R. E., Pääbo, S., 2002a. Intra- and interspecific variation in primate gene expression patterns. *Science* **296** (5566), 340–343.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., Monaco, A. P., Pääbo, S., 2002b. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**, 869–872.
- Eng, L., Coutinho, G., Nahas, S., Yeo, G., Tanouye, R., Babaei, M., Drk, T., Burge, C., Gatti, R. A., 2004. Nonclassical splicing mutations in the coding and noncoding regions of the ATM Gene: maximum entropy estimates of splice junction strengths. *Hum Mutat* **23** (1), 67–76.
- Fisher, S. E., Vargha-Khadem, F., Watkins, K. E., Monaco, K. E., Pembrey, M. E., 1998. Localisation of a gene implicated in a severe speech and language disorder. *Nat Gen* **18**, 168–170.
- Freeland, S. J., Knight, R. D., Landweber, L. F., Hurst, L. D., 2000. Early fixation of an optimal genetic code. *Mol Biol Evol* **17** (4), 511–518.
- Freeman, W. J., 2003. Evidence from human scalp electroencephalograms of global chaotic itinerancy. *Chaos* **13** (3), 1067–77.
- Furlanello, C., Serafini, M., Merler, S., Jurman, G., 2003. Entropy-based gene ranking without selection bias for the predictive classification of microarray data. *BMC Bioinformatics* **4** (1), 54.
- Haken, H., 2002. *Brain dynamics. Synchronization and Activity Patterns in Pulse-Coupled Neural Nets with Delays and Noise*. Springer, New York.
- He, X., Zhang, J., 2005. Gene complexity and gene duplicability. *Curr Biol* **15** (11), 1016–21.
- Knee, R., Murphy, P. R., 1997. Regulation of gene expression by natural antisense RNA transcripts. *Neurochem Int* **31** (3), 379–392.
- Kumar, M., Carmichael, G. G., 1998. Antisense RNA: Function and fate of duplex RNA in cells of higher eukaryotes. *Microbiol Mol Biol Rev* **62**, 1415–1434.
- Lai, C. S. L., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., Monaco, A. P., 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523.

- Lipman, D. J., 1997. Making (anti)sense of non-coding sequence conservation. *Nucleic Acids Res* **25** (18), 3580–3583.
- Mc Kusick, V. A., 1992. *Mendelian inheritance in man. Catalogues of autosomal dominant, autosomal recessive, and x-linked phenotypes*. Vol. 10. John Hopkins University Press, Baltimore.
- Nakamura, Y., Gojobori, T., Ikemura, T., 1999. Codon usage tabulated from the international DNA sequence databases; its status 1999. *Nucleic Acids Res* **27** (1), 292.
- Oresic, M., Shalloway, D., 1998. Specific correlations between relative synonymous codon usage and protein secondary structure. *Mol Biol* **281** (1), 31–48.
- Pinker, S., 2001. Talk of genetics and vice versa. *Nature* **413** (6855), 465–466.
- Quiroga, R. Q., Arnhold, J., Lehnertz, K., Grassberger, P., 2000. Kulback-Leibler and renormalized entropies: applications to electroencephalograms of epilepsy patients. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* **62** (6 Pt B), 8380–6.
- Richard, I., Beckmann, J. S., 1995. How neutral are synonymous codon mutations? *Nat Gen* **10**, 259.
- Schreiber, T., 1999. Interdisciplinary application of nonlinear time series methods. *Physics Report*, 1–64.
- Schug, J., Schuller, W.-P., Kappen, C., Salbaum, J. M., Bucan, M., Stoeckert, C. J., 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol* **6** (4), R33.
- Selkoe, D. J., 1996. Amyloid β -protein and the genetics of alzheimer's disease. *J Biol Chem* **271**, 18295–18298.
- Sharp, P. M., Averof, M., Lloyd, A. T., Matassi, G., Peden, J. F., 1995. DNA sequence evolution: the sounds of silence. *Phil Trans R Soc Lond B Biol Sci* **349**, 241–247.
- Thompson, J. D., Higgins, D. G., Gibson, T. J., 1994. ClustalW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.
- Tyson, J. J., Mackey, M. C., 2001. Molecular, metabolic, and genetic control: An introduction. *Chaos* **11** (1), 81–83.
- Urrutia, A. O., Hurst, L. D., 2001. Codon usage bias covaries with expression breadth and the rate of synonymous evolution in humans, but this is not evidence for selection. *Genetics* **159**, 1191–1199.
- Vinogradov, A. E., 2004. Evolution of genome size: multilevel selection, mutation bias or dynamical chaos? *Curr Opin Genet Dev* **14** (6), 620–6.
- Vogelstein, B., Kinzler, K. W., 1994. Tumor-suppressor genes. X-rays strike p53 again. *Nature* **370**, 174–175.

Wagner, E. G., Simons, R. W., 1994. Antisense RNA control in bacteria, phages, and plasmids. *Ann Rev Microbiol* **48**, 713–742.

Xia, X., 1996. Maximizing transcription efficiency causes codon usage bias. *Genetics* **144** (3), 1309–1320.

Yang, J., Lusk, R., Li, W.-H., 2003. Organismal complexity, protein complexity, and gene duplicability. *Proc Natl Acad Sci U S A* **100** (26), 15661–5.

Yelin, R., Dahary, D., Sorek, R., Levanon, E. Y., Goldstein, O., Shoshan, A., Diber, A., Biton, S., Tamir, Y., Khosravi, R., Nemzer, S., Pinner, E., Walach, S., Bernstein, J., Savitsky, K., Rotman, G., 2003. Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol* **21** (4), 379–86.