COMMENTARY

Comparative genomics: beyond the horizon of the next research grant

Frans Schuit[1]

Abstract With the development of agriculture and food processing techniques, humanity has recently challenged the rules of a billion-year-old experiment called evolution. In this experiment the availability of food in a particular niche has been one of the major driving forces to shape particular species. Comparative genomics is a new research discipline that investigates two or more genomes from different species in order to find specific genetic adaptations that explain a 'workable match' between genetic make-up and environmental constraints such as nutrition. Three recent examples in the literature illustrate how selection of particular genes can contribute to species-specific adaptations that allow them to recognise, secure and digest particular types of food and metabolise its ingredients. There is growing consensus that the recent changes in human diet and physical activity play an active role in the rapid growth of the prevalence of obesity and diabetes. The working hypothesis of the present article is that in the future a more advanced level of comparative genomics of the many natural workable matches of natural species will lead to a much better understanding of the dynamics and regulation of integrated metabolism. It is anticipated that this deeper understanding will lead to novel insights into the mechanism of human diabetes and new strategies for diabetes prevention and treatment. This is one of a series of commentaries under the banner '50 years forward', giving personal opinions on future perspectives in diabetes, to celebrate the 50th anniversary of *Diabetologia* (1965–2015).

Different timescales

When submitting a 5 year research proposal, scientists are aware that the experiments planned for the end of the 5 years are subject to uncertainty. Working hypotheses can metamorphose under the influence of new data, and technological innovations that were unheard of yesterday can be common practice tomorrow. An accurate prediction of what diabetes research will focus on in 2065 therefore seems impossible. Yet, at the risk of being associated with the practice of soothsayers, I have taken the opportunity, on this occasion, the 50th anniversary of *Diabetologia*, to express some hopes and fears about research ideas that lie hidden beyond the horizon of the next research funding opportunity.

As a starting point, during the past 50 years the prevalence of diabetes increased substantially, first in the industrialised part of the world, later in other countries [1]. Although the exact mechanism of this growth is still debated, there is consensus about the idea that an accelerating series of cultural changes rapidly changed our lifestyle (diet, lack of exercise) in such a way that it somehow causes a mismatch between our genes (which evolve very slowly) and the novel environment. Indeed, the timescale of evolution is very different from the short amount of time required for the development of novel science and technology. Agriculture was developed about ten thousand years ago, and most techniques that are relevant to

✉ Frans Schuit
   Frans.Schuit@med.kuleuven.be

[1] Department of Cellular and Molecular Medicine, KU Leuven, Herestraat 49, B3000 Leuven, Belgium

the types of food we eat today were invented much more recently. This is in contrast to the evolutionary distance (5 million years ago [5 mya]) to the point of the last common ancestor of our species and our most closely related family members, the two African chimpanzee species. The last mass extinction event (during the Cretaceous–Tertiary boundary, about 66 mya) dramatically reduced the biodiversity at that moment. In its direct aftermath, biodiversity exploded again: modern birds survived as the sole descendants of the theropod dinosaurs [2]; this expansion occurred in parallel with the radiation of current mammalian taxa.

At present, our planet hosts about 10,000 different bird species and approximately 5000 different mammalian species. What defines these species is today mainly described in terms of anatomy (e.g. shape of bones, colour of fur or feathers) and natural habitat. But with the rapid advances in DNA sequencing technology, we are beginning to gain a deeper understanding of what makes a species unique and how its genetic make-up matches the challenge imposed by the environmental conditions. With the risk of oversimplifying, this problem could be narrowed down to metabolism, in that the genetic make-up of each species has evolved to a workable match between the metabolic needs during a lifetime with all the required flexibility and the environmental constraints of available food. Imagine if we could understand the general principles and essential details of the 15,000 specific matches of individual bird and mammalian species through a deep understanding of general and specific metabolic demands and the nutritional constraints imposed by the environment. Could we connect that vast amount of information to understand the origin of a mismatch such as occurs when we develop diabetes, obesity or the metabolic syndrome? A hope for the future is that a deeper understanding of the plethora of workable matches in nature could inspire innovative ideas for the prevention of diabetes, obesity or the metabolic syndrome in humans. Whether or not the level of biodiversity will have drastically decreased by 2065 depends on whether or not our thriving species, with more than 7.2 billion individuals at this moment, will succeed in finding sustainable solutions for the habitats and ecosystems of other species on this planet [3].

All animals are equal, but some animals are more equal than others

The subtitle refers of course to George Orwell's *Animal Farm*. Perhaps the phrase can also be applied (with a different meaning) to metabolism and its regulation. Vertebrates use glucose, NEFA and ketone bodies as important substrates for ATP synthesis, a process that is critical for sustaining life. Indeed, the need for new ATP molecules imposes itself every second on every living cell, yet for most animals the availability of food that contains the substrate for ATP synthesis often depends on the opportunity of the day. Metabolism has therefore acquired powerful regulatory principles that ensure a constant supply of substrate for energy metabolism. Two of these principles are glucose homeostasis and energy balance, and the tissues and molecules involved are highly conserved (Fig. 1). Consumption of food that contains carbohydrates causes a rise in plasma glucose and this will trigger pancreatic beta cells to release insulin. The physiological need for extra circulating insulin is 'sensed' by rapid uptake of glucose into beta cells via glucose transporters. The rise in circulating insulin is detected by insulin receptors on liver, skeletal muscle and adipose tissue cells. This recognition initiates a series of events that accelerate glucose uptake and metabolic glucose disposal. In skeletal muscle and fat of mammals insulin-mediated glucose uptake is ensured by the glucose transporter isoform GLUT4. Failure of beta cells to respond with sufficient insulin release or resistance of target organs to insulin causes elevated blood glucose and—when more severe—diabetes. At the same time, insulin also regulates the process by which dietary fat is distributed towards triacylglycerol deposits in white adipose tissue, the main energy reserve in vertebrates. In the fasted state, substrate from endogenous reserves is mobilised at rates needed to sustain ATP production. Pancreatic glucagon and adrenaline (epinephrine) are key signalling molecules during this state and regulate glycogen breakdown and the metabolic conversion of fatty acids into ketone bodies in the liver and lipolysis in fat tissue.

Although the overall scheme of glucose homeostasis and energy balance is conserved in higher animals, there are remarkable differences between species. First of all, the strategy of each species to maximise each day's opportunities for finding the food that is needed, varies enormously. Second, the genes that operate in glucose homeostasis and energy homeostasis seem to vary among different species, and it is possible that these differences arose as a consequence of different feeding strategies. One example is that humans have only one insulin gene, but in rodents a second insulin gene (*Ins1*) was 'copied' as a processed retrotransposon and integrated far from the original gene. This happened about 20 mya in the ancestor of mice and rats and it could have contributed to natural selection by dietary constraints [4]. Another example of a difference between rodent and human beta cells relates to the above-mentioned uptake of glucose by beta cells. In rats this occurs via the GLUT2 isoform, but it seems that human beta cells primarily use GLUT1 for this purpose [5]. Recent genome-wide analysis of gene expression in mouse and human beta cells indicates that there are many differences between the two species [6] and we currently only have a rudimentary understanding of why these differences are present.

There appears to be an even more dramatic difference in glucose homeostasis between mammals and birds. The typical fasting glucose level of a healthy bird is above 10 mmol/l, which would be called diabetes by mammalian standards.
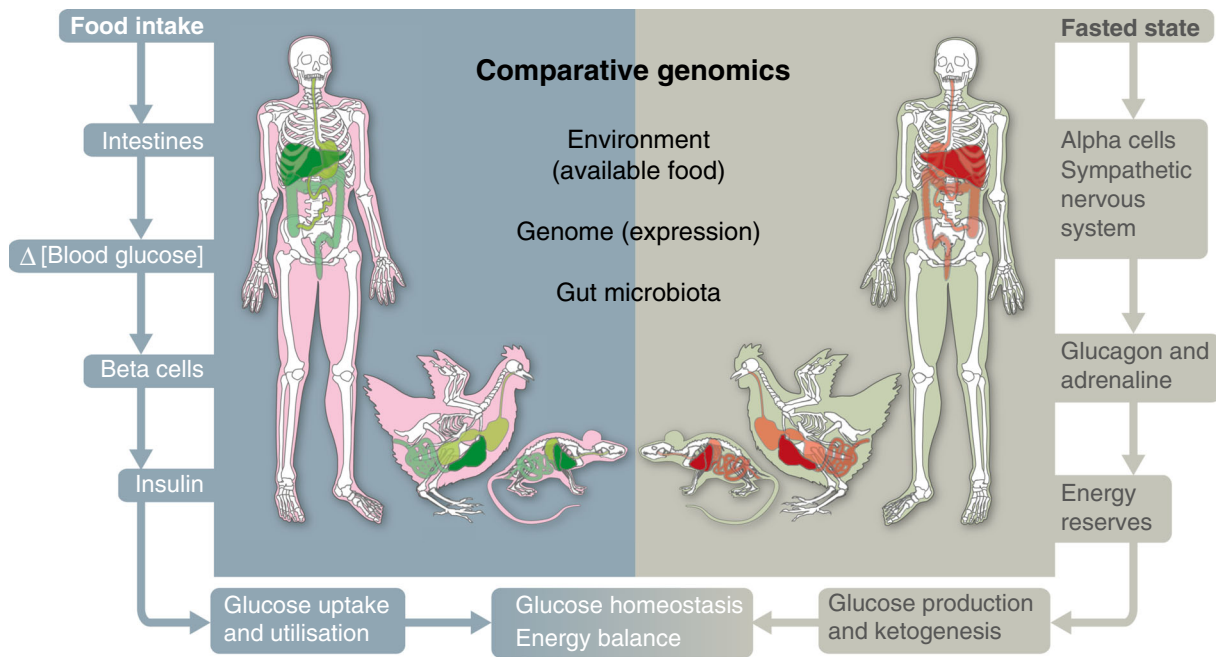
Fig. 1 Comparative genomics in the context of metabolism and metabolic regulation in different species that evolved towards a successful match between environmental conditions (diet, feeding pattern) and the enzymes and regulators encoded in the species' genome and the metagenome of its gut microbiota [17]. The flowcharts around the central image show some of the important elements that contribute to glucose homeostasis and energy balance during the fed and fasted state (see text for explanation). Nutrients other than glucose (butyrate in ruminants, amino acids in carnivores) may be important regulators of insulin release. Comparative genomics can integrate genome information, gut microbiota and the available food supply in the ecological niche to gain an understanding of the workable match between metabolism and diet in different vertebrate species

What is responsible for this elevated blood glucose is an interesting question. On the one hand, birds are very insulin resistant and seem to lack the GLUT4 transporter in skeletal muscle. In the most recent of a series of observations, Welch Jr et al [7] failed to detect GLUT4 mRNA and protein in muscle of zebra finch and hummingbird (the latter has circulating glucose levels of 30–40 mmol/l during feeding on nectar; see below). On the other hand, the avian endocrine pancreas contains more alpha cells than beta cells [8]. How the chronic hyperglycaemia in birds is explained (absence of GLUT4 or high alpha/beta cell ratio) and how chronic hyperglycaemia is tolerated without the development of diabetes complications seem interesting directions for future research. A new discipline in the biological sciences, explained in the next section, may be of help.

Comparative genomics

Comparative genomics is the field of biology in which DNA sequence information from genomes of different life forms is compared. The idea is that differences in genome sequence (genotype) contribute to differences in genome function and therefore explain differences between phenotypic traits. The field started two decades ago with sequence comparisons of bacterial genomes, which are several orders of magnitude smaller than vertebrate genomes. However, with the advances in sequencing technology and data analysis, the same principle can be applied to the complex genomes of higher animals [9], which typically contain a few gigabase pairs per haploid cell. This approach allowed the comparison of DNA sequences of closely related species in order to answer questions related to the unique phenotypic traits of each species. A recent example [10] is the comparative genomics of 120 individual birds selected from roughly a dozen species of 'Darwin's finches' that live on the Galapagos Islands and radiated from a common ancestor during the last million years. Darwin was the first to propose that the match between beak morphology of a bird species and its available food sources on a given island was the work of natural selection, and these ideas have inspired later generations of ornithologists [11]. The recently performed comparative DNA analysis [10] proposes a prominent role for a 250 kilobase DNA region near the *ALX1* gene, which is also involved in craniofacial development in humans.

A second comparative genomic analysis that helps to explain the match between genetic make-up and environmental constraints of food intake is the comparison of genome information from 93 brown or polar bears [12], two closely related species that radiated quite recently (between 0.3 and 0.5 mya) from common ancestors in northern Europe and Canada. The diet of the two current species is different: whereas the brown

bear is a vegetarian, with most of its energy intake in the form of carbohydrates, the diet of polar bears consists of seals, which are rich in protein, triacylglycerol and cholesterol. Circulating LDL-cholesterol levels in polar bears are indeed much higher than those in brown bears, and an interesting observation of the comparative genomic study was that the *APOB* gene, which encodes the apoprotein that addresses LDL-cholesterol towards LDL receptors, has been under selective evolutionary pressure in polar bears [12]. These two examples of analyses between closely related species illustrate the power of this approach to correlate genotypic information with phenotypic specialisation. In each case the environmental pressure (type and abundance of, and influence of season on, available food) could have been the primary evolutionary driving force for the selection of genetic changes in particular genes so that related species living today evolved towards a different match between genetic make-up and the environmental niche.

The same principle can be used to compare genes of more distantly related species, such as birds and mammals. One example is the comparative genomics of umami and sweet taste receptors, which detect two of the five main taste qualities. Each of these receptors work as protein heterodimers that are encoded by a pair of taste receptor genes: TAS1R1–TAS1R3 for umami (detecting amino acids such as glutamate, alanine and serine) and TAS1R2–TAS1R3 for sweet taste (detecting fructose, sucrose and glucose). Among mammals, several carnivorous species, such as cats, hyenas, otters and seals independently have lost their functional *TAS1R2* gene by different pseudogenisation events, such as nonsense mutations that create a premature stop codon [13]. In line with the mammalian carnivores, all modern birds (descendants of carnivorous theropod dinosaurs) have completely lost the *TAS1R2* gene. The question then arises of how hummingbirds (which feed on sweet-tasting nectar) can sense the quality of their food. The intriguing answer is that the ancestral *TAS1R1* gene in hummingbirds was specifically reprogrammed in order to 'reinvent' the lost sweet taste receptor [14]. The mechanism was found by comparing the hummingbird *TAS1R1* gene to the chicken or swift orthologues. Specific mutations in hummingbird *TAS1R1* exons occurred uniquely in a handful of codons that encode protein residues needed for the ligand-binding properties of the receptor. The result is a reprogrammed hummingbird TAS1R1–TAS1R3 system that favours sucrose, fructose and the sweetener sucralose over alanine and serine [14]. The umami receptor was hereby transformed into a sweet taste receptor. This example illustrates how selective gene mutations can help to optimise the strategy of a species to find available food in its ecological niche.

Beyond the horizon of the next research grant

The three examples discussed in the previous section indicate that a comparative genomic approach can provide new information on the mechanisms via which genomes evolved towards a workable match between encoded proteins and the natural environment. Recent research indicates that this approach can be scaled up considerably. At the time of writing, the genomes of more than 75 mammalian species and 51 different birds are available in public databases (www.ncbi.nlm.nih.gov/gene/ accessed 9 April 2015). At the end of 2014, a series of articles was published that presented the results of a comparative genome analysis of a large dataset including the genomes of 48 bird species, and one of those studies traced the evolutionary history of the explosive radiation of most current species [2]. This year we also celebrate the 50th anniversary of Gordon Moore's prediction that the overall processing power of computers doubles every few years. Although the rate of growth of computational power has tended to decline a bit over the past years, there is little reason to doubt further growth during the next decades. Thus, we can anticipate significant up-scaling of the capacity for DNA analysis and the power to mine datasets of gigantic proportions. Understanding and preserving biodiversity could be the prime driving force behind sequencing the genomes of many hundreds of other vertebrate species. This effort would gain value if it led to a deeper understanding of how the inherited DNA sequence information contributes to fitness in a competitive world. For example, from a metabolic point of view, it seems relevant to understand how genes that encode enzymes or metabolic regulatory proteins contribute to the adaptation of a given species to the food it needs to survive. We are largely lacking this type of understanding today. Instead, we use model organisms such as genetically modified mice, which sometimes have their own artefacts [15], in an attempt to understand the underlying mechanisms in human diabetes. We cannot ignore the fact that these model organisms evolved over millions of years towards a specific match between metabolism and diet. It would be wrong to state that every species succeeded in finding a perfect match, and evolution could be seen as an experiment 'in progress' with many failures. In a world that rapidly changes, e.g. by global warming, the reasonable match of yesterday may rapidly shift into a lethal mismatch of tomorrow, and extinction events could be a possible outcome when genes have no time to mutate. Another limitation of the current working hypothesis is that other major environmental factors, such as defence against pathogens may shape the evolution of a species and its metabolism, as illustrated by mutations in the gene encoding human glucose-6-phosphate dehydrogenase in areas where malaria is endemic [16]. Finally, it should be noted that some differences between genomes were established by random drift rather than by positive selection. Nevertheless, it

seems reasonable to underline our lack of understanding of how different species match the sequence of their genes to the challenge of nutrient detection and integrated metabolism imposed by a certain type of diet. Much fundamental research is needed for a deeper level of understanding that may bring us to new directions currently hidden by the horizon. When the enormous dataset of natural DNA variation between species is linked to the specific DNA mutations that are linked to human diabetes we may expect novel approaches to understand diabetes, in order to cure tomorrow's patients and prevent disease in others.

Duality of interest The author declares that there is no duality of interest associated with this manuscript.

Contribution statement The author was the sole contributor to this paper.

References

1. Zimmet PZ, Magliano DJ, Herman WH, Shaw JE (2014) Diabetes: a 21st century challenge. Lancet Diabetes Endocrinol 2:56–64
2. Jarvis ED, Mirarab S, Aberer AJ et al (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331
3. Carroll SP, Jørgensen PS, Kinnison MT et al (2014) Applying evolutionary biology to address global challenges. Science 346:313
4. Shiao MS, Liao BY, Long M, Yu HT (2008) Adaptive evolution of the insulin two-gene system in mouse. Genetics 178:1683–1691
5. De Vos A, Heimberg H, Quartier E et al (1995) Human and rat beta cells differ in glucose transporter but not in glucokinase gene expression. J Clin Invest 96:2489–2495
6. Benner C, van der Meulen T, Cacéres E, Tigyi K, Donaldson CJ, Huising MO (2014) The transcriptional landscape of mouse beta cells compared to human beta cells reveals notable species differences in long non-coding RNA and protein-coding gene expression. BMC Genomics 15:620
7. Welch KC Jr, Allalou A, Sehgal P, Cheng J, Ashok A (2013) Glucose transporter expression in an avian nectarivore: the ruby-throated hummingbird (*Archilochus colubris*). PLoS One 8:e77003
8. Bonner-Weir S, Weir GC (1979) The organization of the endocrine pancreas: a hypothetical unifying view of the phylogenetic differences. Gen Comp Endocrinol 38:28–37
9. Howe K, Clark MD, Torroja CF et al (2013) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503
10. Lamichhaney S, Berglund J, Almén MS et al (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375
11. Lack D (1947) Darwin's finches. Cambridge University Press; reissued (1961) Harper, New York
12. Liu S, Lorenzen ED, Fumagalli M et al (2014) Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell 157:785–794
13. Jiang P, Josue J, Li X et al (2012) Major taste loss in carnivorous mammals. Proc Natl Acad Sci U S A 109:4956–4961
14. Baldwin MW, Toda Y, Nakagita T et al (2014) Sensory biology. Evolution of sweet taste perception in hummingbirds by transformation of the ancestral umami receptor. Science 345:929–933
15. Brouwers B, de Faudeur G, Osipovich AB et al (2014) Impaired islet function in commonly used transgenic mouse lines due to human growth hormone minigene expression. Cell Metab 20:979–990
16. Taylor SM, Fairhurst RM (2014) Malaria parasites and red cell variants: when a house is not a home. Curr Opin Hematol 21:193–200
17. David LA, Maurice CF, Carmody RN et al (2014) Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563