Comparison of Ataxia Telangiectasia Mutated's nucleotide sequence in different model organisms

Jaspreet K. Rishi

Introduction

Ovary is the female gonad that houses the female gamete, the egg. Females are born with a finite number of oocytes (primordial follicles) which once expended, cannot be replenished. The mammalian ovary has oocytes present at different levels of development at any given time point. When a female hits puberty, folliculogenesis begins every month where one primordial follicle fully matures to undergo ovulation and the egg is released. In case the egg is fertilized, the remaining follicular tissue regresses into corpus luteum which releases gonadotropins, including progesterone to help maintain pregnancy[1]. Depletion of oocytes is a natural process and once a female is ceased off all the oocytes, she undergoes menopause. Menopause is marked by low levels of estrogen which may increase the risk for certain medical conditions such as osteoporosis, colon and ovarian cancer, periodontal disease, tooth loss and cataract formation[4]. When menopause happens before the age of 40, it is termed as premature ovarian insufficiency and it can be caused by a variety of factors, including environmental toxins and lifestyle factors, leading to infertility. Obesity and the environmental toxin DMBA, 7,12 Dimethylbenz[a]anthracene, are two such factors that cause detrimental effects on the oocytes. Obesity is a rising concern in the United States. Obesity causes adverse effects on reproduction including reduced conception and implantation[2], impaired fecundity[3] and infertility[2]. Obesity can induce early menarche in young girls and early menarche has been implicated in causing premature menopause[5]. Previous studies from our labs have shown that ovaries from obese females have a greater level of DNA damage as well[8].

Polycyclic aromatic hydrocarbons are produced by burning organic matter and are present ubiquitously. PAH's were also recently listed in the top 10 priority concern chemicals by the Agency for Toxic Substances and Disease registry[6]. DMBA, produced by burning organic

matter, through wildfires, waste incineration , during coal tar and coke production, smoking etc. is one such polycyclic aromatic hydrocarbon that causes ovarian toxicity. Ovotoxicants can cause depletion in oocytes housed within growing follicles, causing temporary infertility and also cause demise in primordial follicles leading to permanent infertility. DMBA is ubiquitous and once ingested, it is bioactivated by biotransformation enzymes to a genotoxicant epoxide form in the ovary[7]. PAHs, including DMBA have detrimental effects on fetal development and gestational exposure to PAHs are associated with neural tube defects[10], behavioral issues[9], gastroschisis[10], cleft lip[11], preterm birth[12], reduced birth weight[13] and increased incidence of metabolic syndrome[14].

Ovarian toxicology affects epigenetic regulation as well. Post-translational histone modifications impact female fertility via regulating follicle formation and maturation. Histone modifications involved in the DNA damage repair response do to ovotoxicity include phosphorylation of H2AX, by ATM (Ataxia Telangiectasia Mutated)[15], resulting in the recruitment of downstream DNA damage repair proteins as shown in fig1. ATM mediates cell cycle arrest and is one of the deciding proteins for the fate of cell when it suffers DNA damage (20). Depending on the extent of damage, the cell will either undergo DNA damage repair or apoptosis. ATM has also shown to control recombinational repair of DNA Damage[16] thus it is an important target to look at when learning about DNA damage repair.

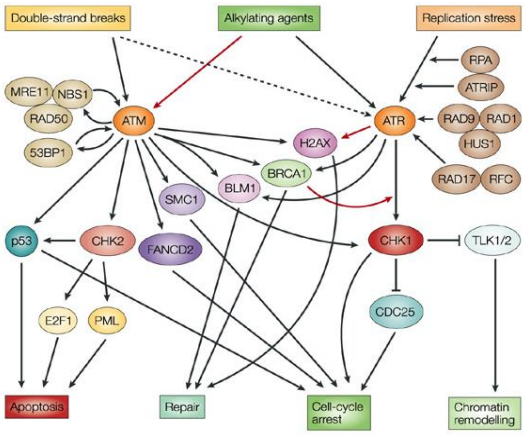

Fig 1. DNA damage repair pathway[17].

DNA damage is a big contributor to infertility and DDR due to ovotoxicity is under-explored. Our aim is to explore DDR in mice ovaries and deduce the mechanisms by which ATM's role is affected by toxins and obesity, thereby causing infertility. In order to determine whether the results inferred from studying the ATM protein in mice will be translatable to humans, we generated phylogenetic trees of the ATM protein in different species. Protein sequences from a total of 10 highly used model organisms were used to generate the best phylogenetic tree using maximum liklehood, to find how similar the sequence of ATM is in mice and humans and which organism has the most similar ATM sequence with that of human ATM.

Materials and Methods

The nucleotide sequences of the protein ATM from 10 species were collected from NCBI. In order to remove contamination, only nucleotide sequences of the coding region of the gene were used. MAFFT and RAxML-NG tool was used phylogenetics analyses. The model used was GTR+R. Bootstrap analysis were run for 1000 replicates or till it converged. The ITOL website was used to visualize some trees. Mr. Bayes package was used to run Bayesian analysis for hypothesis testing.

Results

Maximum likelihood

Since there were more than 2 sequences to align, a multiple sequence alignment tool was used to look at the similarities among multiple sequences. The fasta files from 10 species were aligned locally and a maximum likelihood tree was generated using MAFFT, as shown in figure 1. The seed value was set as 100 to have a high reproducibility.

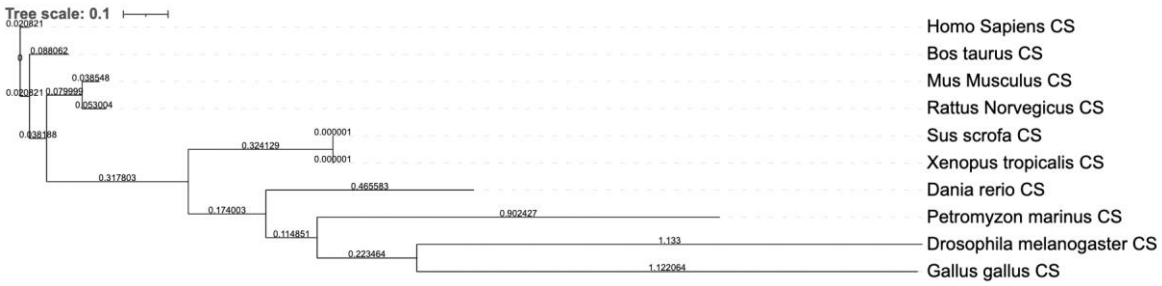

Figure 2. Best tree produced by maximum likelihood.

Bootstrap analysis

Non-parametric bootstrap analysis was run with *Homo sapiens* as the outgroup. The default for a 1000 replicates was chosen. Bootstrap support function was used to analyze confidence of each branch generated by the maximum likelihood tree. The tree thus generated is provided in fig 2.

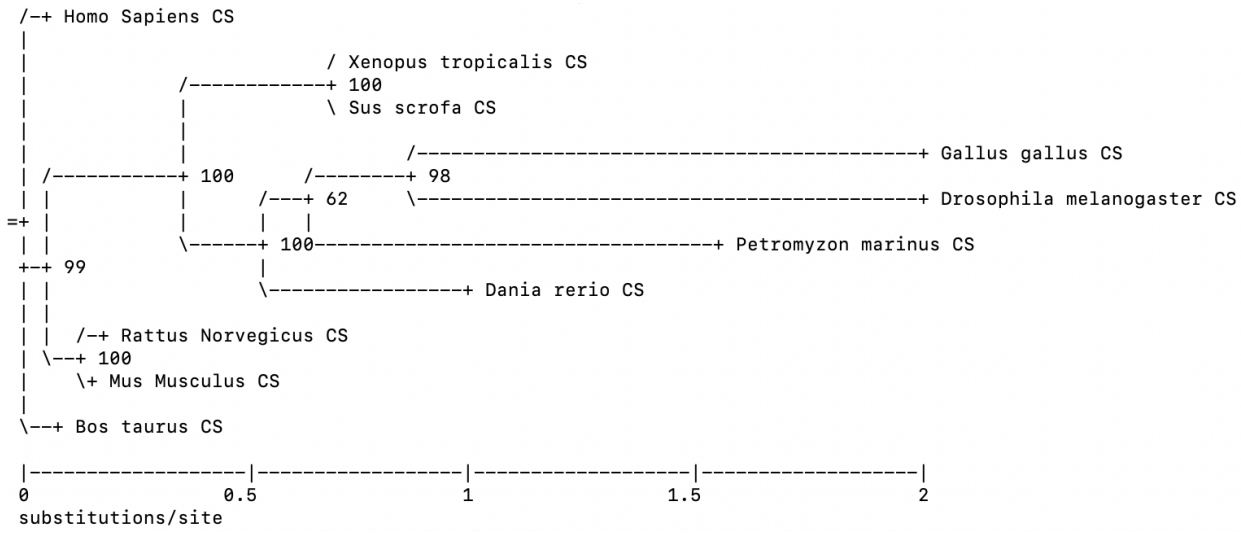

Fig 3. Bootstrap support tree produced by 1000 replicates

Hypothesis Testing using Bayesian analysis

The hypothesis that Homo sapiens and Mus musculus are closely related was tested by Bayesian analysis and the results from the positive constraint and negative constraint are given in tables 1 and 2.

Run	Arithmetic mean	Harmonic mean
1	−35450.26	−35460.93
2	−35449.98	−35462.71
TOTAL	−35450.11	−35462.17

Table1. Mean values of positive constraint i.e. *Homo sapiens* and *Mus musculus* are closely related.

Run	Arithmetic mean	Harmonic mean
1	−34911.73	−34922.34
2	−34911.89	−34923.14
TOTAL	−34911.80	−34922.82

Table 2. Mean values of negative constraint i.e. *Homo sapiens* and *Mus musculus* are not closely related.

Discussion

The alignment was first generated by using the entire gene sequence that codes for the ATM protein from different species which contained a huge number of gaps. In order to remove the introns, just the coding sequence was used, and the sequences aligned better, although there were still quite a number of gaps that can be attributed to huge number of differences found in eukaryotes. The maximum likelihood tree was generated, and it was rooted by using *Homo sapiens* as an outgroup as that's the species we are trying to compare the rest of the species sequences with. Outgroup rooting also allowed to determine the relatedness of sequences aligned. The best tree showed that the ATM protein's sequence in humans was the most closely

related to that of *Bos Taurus* (cattle) as the branch length of the Bos taurus was the shortest at 0.088062. Interestingly, the next closest leaf of the tree was *Mus musculus* (mouse) at a length of 0.138868. The next leaf was *Rattus norvegicus*(rat), followed by *Sus scrofa*(pig) and so on. The phylogenetic tree generated was not surprising but rather expected as it was reflective of what is also seen when the entire genomes of these species are used to generate a phylogenetic tree[18]. Had it been different, it would be indicative of lesser or greater differences in a particular gene sequence, in this case ATM, as a ratio of differences in the entire genome. Bootstrap analysis support showed high values, some even 100, showing high confidence level in the branches of the tree generated. Bayesian analysis was performed for hypothesis testing. Default value of 2 chains was used for mcmc. Positive constraint was set as *Homo sapiens'* and *Mus musuculus'* ATM sequence being closely related and the negative constraint was that two sequences were not closely related. The log likelihood values generated after setting the two constraints differed each other by 539.35. Using the formula, $K = \ln[BF(M0,M1)] = \ln[P(X \mid M0)] - \ln[P(X \mid M1)]$, $K = 2.29$ and $BF = 9.87$. Since $K > 1$; we have support for disregarding the negative constraint and stating that the first model, i.e. ATM sequence of *Homo sapiens* and *Mus Musculus* are similar.

Since experimental genetics studies in humans are highly restricted due to ethical reasons, model organisms must be used and it is important to pick a model organism which is closely related to humans if the results are to be translated to humans. For this purpose, it is also important to not just look at the phylogenies of the entire genomes of the model organisms, but to dive into the specifics and focus on similarities among the gene sequences. Similarities in the coding sequences of the gene represents functional equivalence and evolutionary relationships between the protein among different species. *Bos taurus, Mus musculus, Rattus norvegicus* and *Sus scrofa* are the species mostly found in literature pertaining to reproductive studies. The ATM gene sequence of 3 out of these 4 species are the most closely reated. In our lab, Mus musculus is used as the model organism and the phylogenetic pattern seen here adds

credibility to our studies and provides confidence to be able to hypothesize that the DNA Damage repair pathway that ATM is involved in in *Mus musculus* is similar to what one might expect in *Homo sapiens*.

Conclusions

Analyzing the sequence of the protein that we are learning about in *Mus musculus* and comparing it other species sequences, especially that of *Homo sapiens* and finding its relatedness adds to the soundness of our study and makes it more translatable to humans. This alone is not enough evidence to claim that any results obtained about this protein in *Mus musculus* can now be applied to the function of ATM in Homo sapiens. But until we are able to perform studies directly in *Homo sapiens*, this result takes us one step closer to making that claim.

References

1) Niswender, Gordon D., et al. "Mechanisms controlling the function and life span of the corpus luteum." *Physiological reviews* 80.1 (2000): 1-29.
2) Brewer CJ, Balen AH. The adverse effects of obesity on conception and implantation. Reproduction. 2010;140(3):347-364.
3) Kulie T, Slattengren A, Redmer J, Counts H, Eglash A, Schrager S. Obesity and women's health: an evidence-based review. J Am Board Fam Med. 2011;24(1):75-85.
4) Traci C. Johnson. “Premature Menopause.” *WebMD.* June 14, 2020. www.webmd.com/menopause/guide/premature-menopause-symptoms
5) Wang Z, Dang S, Xing Y, Li Q, Yan H. Correlation of body mass index levels with menarche in adolescent girls in Shaanxi, China: a cross sectional study. BMC Womens Health. 2016;16:61.
6) Agency for Toxic Substances and Disease Registry. Priority List of Hazardous Substances. In: Centers for Disease Control, ed2013.

7) Ganesan S, Nteeba J, Keating AF. Enhanced susceptibility of ovaries from obese mice to 7,12-dimethylbenz[a]anthracene-induced DNA damage. Toxicology and applied pharmacology. 2014;281(2):203-210
8) Perera FP, Tang D, Rauh V, et al. Relationships among polycyclic aromatic hydrocarbon-DNA adducts, proximity to the World Trade Center, and effects on fetal growth. Environmental health perspectives. 2005;113(8):1062-1067.
9) Perera FP, Wang S, Vishnevetsky J, et al. Polycyclic aromatic hydrocarbons-aromatic DNA adducts in cord blood and behavior scores in New York city children. Environmental health perspectives. 2011;119(8):1176-1181.
10) Lupo PJ, Langlois PH, Reefhuis J, et al. Maternal occupational exposure to polycyclic aromatic hydrocarbons: effects on gastroschisis among offspring in the National Birth Defects Prevention Study. Environmental health perspectives. 2012;120(6):910-915.
11) Langlois PH, Hoyt AT, Lupo PJ, et al. Maternal occupational exposure to polycyclic aromatic hydrocarbons and risk of oral cleft-affected pregnancies. Cleft Palate Craniofac J. 2013;50(3):337-346.
12) Choi H, Rauh V, Garfinkel R, Tu Y, Perera FP. Prenatal exposure to airborne polycyclic aromatic hydrocarbons and risk of intrauterine growth restriction. Environmental health perspectives. 2008;116(5):658-665.
13) Langlois PH, Hoyt AT, Desrosiers TA, et al. Maternal occupational exposure to polycyclic aromatic hydrocarbons and small for gestational age offspring. Occup Environ Med. 2014;71(8):529-535.
14) Yan Z, Zhang H, Maher C, et al. Prenatal polycyclic aromatic hydrocarbon, adiposity, peroxisome proliferator-activated receptor (PPAR) gamma methylation in offspring, grand-offspring mice. PloS one. 2014;9(10):e110706.

15) Ganesan S, Bhattacharya P, Keating AF. 7,12-Dimethylbenz[a]anthracene exposure induces the DNA repair response in neonatal rat ovaries. Toxicology and applied pharmacology. 2013;272(3):690-696.
16) Morrison, Ciaran, et al. "The controlling role of ATM in homologous recombinational repair of DNA damage." *The EMBO journal* 19.3 (2000): 463-471.
17) Corcoran, Niall M., et al. "Molecular pathways: Targeting DNA repair pathway defects enriched in metastasis." *Clinical Cancer Research* 22.13 (2016): 3132-3137.
18) StemMed Labs. https://www.stemedlabs.org/blog/how-do-biologists-decide-what-organism-to-study/