nature portfolio

Peer Review File

The emergence of Sox and POU transcription factors predates the origins of animal stem cells

Corresponding Author: Dr Ralf Jauch

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
General comments

This report describes the discovery of members of Sox and POU transcription factor families in choanoflagellate and filasterean unicellular holozoans. This pushes the origin of these families back beyond the origin of the metazoans, which has been suggested in previous publications. Members of both families – Sox2 (SoxB) and Oct4 (POU5) specifically - are involved in inducing stemness in vertebrates, and have been shown or implicated to be involved in stemness in other bilaterians and cnidarians.

The evidence for the ancestral holozoan Sox gene having biochemical properties remarkably similar to extant Sox2 is compelling and very interesting. Multiple lines of phylogenetic and experimental evidence support the authors' claim that genes identified in choanoflagellates can substitute for mammalian Sox2 in a range of molecular and cell contexts, including partnering with mammalian Oct4 in DNA-binding analyses using the canonical (mammalian) SoxOct DNA cis-element and inducing mouse PSCs. Given the similarity of the choanoflagellate Sox sequence to metazoan Sox2/SoxB and that these HMG DNA-binding domains were expressed as chimerics that were flanked by mammalian Sox2 C- and N-sequences, it might be expected that these Sox hybrids will be functional in mice.

In a set of similar analyses it was shown that the choanoflagellate POUs do not function as the mammalian Oct4 despite possessing bona fide POU and homeodomains with conserved residues responsible for DNA binding. The phylogenetic tree shown in Fig. 5A differs from previously published trees of the POU family that used only metazoan sequences, with sponge and fly sequences sitting outside the recognised 6 subfamilies, and cladeing within or outside the choanoflagellate POUs. It appears this topology may be "forced" by the parameters set by the authors and described briefly in the Methods (there is no further analyses in the supplementary materials except an alignment of POU and homeodomain sequences, which is convincing in terms of the choanoflagellate sequences being genuine POU HDs but not it terms of their relationship to metazoan POUs). The phylogenetic analyses should be redone and the results reported in detail. Choanoflagellate POUs may not have worked in the mouse system because they are highly divergent.

Even though the authors convincing demonstrated that the ancestral choanozoan (choanoflagellates + animals) Sox gene was likely to have been a member of SoxB/Sox2 subfamily, their interpretation of its role in the evolution of stemness in confusing in places. As written, some readers might infer that the origin of stemness and Sox2 are being conflated. The narrative around the evolution of metazoan pluripotency and potential role of Sox needs to be explained more clearly (see specific comments below). For instance, both Sox and POU genes were detected in only a small number of unicellular holozoans, and it is inferred that these genes were lost in most choanoflagellate and filasterean lineages. If this is the case, is there any aspect about the life cycles of the species these genes were found that might provide insights into role of Sox in these species and thus the role of the ancestral Sox and its potential to be involved in cellular reprogramming or state changes. Related to this point, a discussion on the fact that most of the unicellular holozoan species are represented by a transcriptome(s) and not a genome is warranted. Are these transcriptomes targeted stages of the life cycle or opportunistic (random) sampling? Also, does depth and coverage of sequencing contribute to the inability to find Sox and POU genes in most species?

Specific comments

Abstract

1. Sox genes shows that their molecular ability to induce stemness was already present in the last common ancestor of animals and their unicellular relatives.

This inference does not necessarily follow. Just because an ancestral Sox gene can function in contemporary mammals does not necessary mean that was its ancestral function.

2. Our findings imply that the evolution of stem cells exploited a pre-existing set of transcription factors, where the critical innovation involved an initial change in DNA specificity of POU and the exaptation of the ancestral capacity to interact with Sox transcription factors.

Confusing as written. Is “exploited” the best term to describe this situation? Also, the statement infers that the partnering of SoxB and POU5 was instrumental in the early evolution of stem cells. There is no evidence in this report of these two TFs partnering in unicellular holozoans. Nor is there currently evidence of them partnering in sponges, placozoans, ctenophores or cnidarians, despite the presence of stem cells, SoxB and multiple POUs in these taxa. As such, this statement needs to be tempered and rewritten.

Introduction

1. We propose that the Sox/POU partnership emerged alongside the trajectory towards animals and was likely driven via molecular exaptation of the previously established Sox-POU-DNA dimerization and binding capacities.

See comments above - Abstract 2

Discussion

1. Their molecular features to bind and open up closed chromatin as pioneer factors might have primed Sox proteins for becoming stem cell regulators in animals 4,5,7,31.

Although this is an interesting deduction based on modern Sox-chromatin interactions that seems very plausible, cellular plasticity – allowing cellular responses to exogenous and endogenous signals – likely evolved before a cell state that we recognise as a pluripotent stem cell. Consider reworking this sentence based on these and comments above (Abstract 2).

2. Our data show that many of the biochemical features that render Sox stemness factors are already present in the last common ancestor of animals and choanoflagellates, facilitating its co-option into pluripotency networks (Figures 1E-G,3A-F).

The first part of the sentence is correct but the second half does not follow. As written a pluripotency network was in place without the need for Sox. There seems to contradict other statements in the manuscript that implicate Sox being central to the evolution of pluripotency.

3. As a notable exception, sponges from the demosponge class do not express SoxB in the assumed pluripotent stem cell in this lineage, the archeocytes 55,56, yet this might not extend to other sponge classes.

SoxB is expressed in larval and adult archaeocytes in Amphimedon – see supplementary figures and tables in References 1 and 56.

4. Thus, given the capacity of a choanoflagellate Sox to interact with Oct4 to drive
pluripotency, analogous partnerships might have enabled animal evolution (Figure

Not sure what this means - need to be more explicit. New TF partnerships driving multicellular complexity? For example, bZIPs?

5. Since Sox pre-dates the origin of stem cells, the remaining question is if there was any further molecular innovation required for the establishment of this gene regulatory network in animals.

Again, vague in terms what is meant evolutionary. This seems to be a difficult question to answer given the paucity of GRN information from early-branching animal lineages

Methods

1. We used the human Sox2 protein as query for a BLASTP search against the predicted proteomes of unicellular holozoans, including 22 choanoflagellates (20 transcriptomes and 2 genomes), 4 filastereans, 7 ichthyosporeans …

The lack of Sox and POU genes could have to do with transrcriptome coverage. It would good to see the sequencing data. Are the transcriptomes stage specific?

2. The resulting 756 sequences were aligned using MAFFT, trimmed uing trimAl, and IQTREE to build the phylogeny…

Why isn't this shown? Using POU6 as an outgroup appears to force sponge and othermetazoan POUs outside the recognised metazoan subfamilies. It is unclear why this was done.

3. Protein samples and fluorescently labelled DNA were mixed in EMSA buffer…

Rephrase to make clear these are recombinant proteins, not protein samples from cells.

Reviewer #2

(Remarks to the Author)
The manuscript reports the striking ability of pioneer transcription factors Sox orthologues from unicellular organisms like choanoflagellate and filasterea, to interact with the same DNA motif like the mouse Sox2, and even the ability to induce iPSC formation. This is an exciting finding, and of great interest to a scientific communities interested in chromatin, transcription factors, reprograming, and evolution. The findings shed light on the early evolution of pioneer transcription factors, and pluripotency. There are a couple minor comments that I would like the authors to take care of.

- A question related to the reprogramming experiments with Sox orthologues. I think it would be really important to show how the levels of Endogenous SOX2 are not changing in different reprogramming conditions - when different Sox orthologues are used in the Yamanaka cocktail. A control with a temporary mouseSox2 knockdown and the addition of Yamanaka cocktail should be performed (at least for one Sox2 orthologue). In that case reprogramming should be still observed. Or any alternative experiment to show that the level of endogenous Sox2 is close to zero, and does not change, so that we are sure that the observed differences are clearly attributed to the introduced Sox and not the endogenous.
- Another important control would be to add a chimera with mouse Sox2 HMG and Salhel and/or Phi N and C-termini, and see if that chimera can lead to iPSC (as in Fig. 2A), in order to complete a picture of the role of DBD vs tails in the pluripotency induction. And potentially a control with a Sox2 DBD only (with tails cut off).
- Could the authors also discuss a bit, which part of Sox is the most essential from the evolutionary stand point in terms of pluripotency induction in their opinion, and why? Is it the DBD that interacts with Oct4, or are those the unstructured tails? Based on mutagenesis, one would think it's the DBD, but the chimera of Salhel DBD + native Sox2 tails could not lead to iPSC. Please comment on that.
- Please show the complete DBD sequence of the mouseSox2, Sox17 and the orthologues in the figure1 (instead of S1), and indicate sequence identity scores just for the DBD. The partial (cut out) sequence is misleading. Also please show the complete alignment including the unstructured regions in the supplementary figure 1. Please comment whether there is any sequence similarity to mouseSox2 in the unstructured parts.
- Please comprehensively color code Figure S1 seq alignment, there is no back-and-white requirement.
- Please also discuss whether the sequence similarity between mouseSox2 and different Sox orthologues is in agreement with how "they act": i.e. a higher seq similarity in those orthologues that also induce pluripotency and vice versa.
- Fig. 5B need a complete sequence of the domain, and not a cut out; same in Fig. 4B. Or remover them completely if the sequences are shown later in the supplement.
- Typo on page 4: Holozoa (instead of Holoza)

Reviewer #3

(Remarks to the Author)
Summary
This study investigates the evolutionary origins of pluripotency by examining the functions of two pluripotency genes (Sox2 and Oct4) from organisms outside metazoa. The authors demonstrate the presence of Sox-family HMG box genes and POU family genes in choanoflagellates and show that the Sox genes can bind DNA and activate the expression of Oct4 in mouse embryonic fibroblast cultures. They use this result to suggest that the pluripotency-inducing capacity of Sox genes evolved before the last common ancestor of animals. The approach is bold and creative but I am not convinced that the data they present tell us anything about OSKM-driven pluripotency, which is a vertebrate specific phenomenon. Several concerns are outlined below.

General comments
The main conclusion is that bona fide sox genes were recovered from the two new choanoflagellates (Salpingoeca helianthica and Mylnosiga fluctuans) where only sox-like HMG box genes have been discovered previously (in the two previously sequenced choanoflagellates S. rosetta and Monosiga brevicolis). The authors support this claim using phylogenetics to group the new choanoflagellate sox genes with animal sox genes (to the exclusion of the sox-like genes from S. rosetta and M. brevicolis) and by demonstrating that Salhel sox can replace mouseSox2 in assays of induced pluripotency. From the tree provided (Fig S1) the rationale for including the Salhel and Myflu sequences in the Sox family while relegating the Sros and Mbre sequences to "Sox-like" is not clear. There is very low bootstrap support (~30%) connecting Salhel and Myflu to the animal sox genes and there appear to be some sponge sequences (Scil) that were excluded from the sox clade. Examining the partial alignment provided in Fig S2, it also does not seem there is a clear

reason for excluding the sequence labeled Srosetta-Sox-II from the functional (in vitro) analyses, as it shares all the same “key” residues as the sox genes from human Sox2, Salhel Sox, and MfluSox. If the authors are going to claim they have found something unique to Salhel and Myflu, they would need to perform the same experiments (in mouse embryonic fibroblasts) on (minimally) the Sros-SoxII sequence and demonstrate that it lacks this capacity. Additionally they need to provide better support for the topology they have chosen to represent with their Sox tree by performing an approximately unbiased test (see: Shimodera 2002 Syst Biol) and providing intermediate trees that were generated before they decided to prune sequences.

Independent of whether the HMG box genes they identify from Salhel and Myflu are sox or sox-like genes, it’s not clear what we really learn from the in vitro experiments presented here. The authors show that some small number of mouse embryonic fibroblasts express Oct4 when they are transfected with choanoflagellate sox in place of mouse sox2 but there’s no way for the reader to evaluate the significance of these results. As a negative control the authors perform the same experiment with Sox17 which is known to be incapable of inducing pluripotency in this system. Considering Sox17 is also expressed in the early embryo (at/near the same time as Sox2) and is responsible for driving differentiation of the embryonic endoderm, it is reasonable to assume there has been strong diversifying selection to keep the functions of Sox2 and Sox17 separate in mouse. Thus, Sox17 is not really the appropriate negative control. Rather, I imagine that the authors might recover the same low number of Oct4 expressing cells (as they did with Salhel and Myflu sox sequences) if they repeated these experiments with just about any soxB gene from any invertebrate in place of mouse Sox2. Indeed, their ancestral state analysis (Fig 4) suggests this should be true. (Although, see below for additional concerns about the ancestral state analysis.) The proper control then would be to mis-express the complete (not chimeric) sox ortholog from something like a planarian or a hydra in place of mouse Sox2. If the “pluripotency-inducing” capacity of these sox genes is equivalent to that of the choanoflagellate sox gene, what have we really learned? Neither planarians nor cnidarians use the OSKM gene cassette in their pluripotent stem cells (indeed, neither of these taxa even have a proper ortholog of Oct4 or Nanog) so independent of the behavior of their sox genes, it is not reasonable to imply that the capacity for mouse-like pluripotency is an ancient property of sox genes. What the authors have demonstrated is that the capacity of Sox genes to bind their target DNA sequences is highly conserved across animals, which is very cool, but extending the conclusion to say something about the origin of OSKM-pluripotency is unwarranted. The low-level activation of Oct4 certainly does not derive from an ancient relationship between sox and Oct4 but rather seems like it might be noise (there is even low-level GFP expression in the SoxCEF experiment). The authors would need to examine the capacity of choanoflagellate sox to interact with other pou genes to suggest there was an ancient relationship between these factors.

The title of the manuscript implies the authors have identified an ortholog of Oct4 in choanoflagellates, as other POU genes do not induce pluripotency (see: Nakagawa et al., 2008 Nat Biotech; Jin et al., 2016 Scientific Reports) but this is misleading. Oct4 is a POU5 ortholog and this clade of POU genes seem to be a vertebrate (or perhaps a deuterostome) innovation (Sukparangsi et al., 2022 Nat Comms; Gold et al., 2014 MBE). Importantly, the manuscript lacks a clear definition of how pluripotency is assayed in this system. In some places the authors imply that activation of Oct4-GFP is sufficient to conclude the pluripotent state has been reached, in other places they seem to suggest the expression of other factors is required (e.g., expression of Nanog and Esrrb are shown in Fig2 and S5 but are never introduced or explained anywhere in the text of the manuscript). Coincident with this, the authors imply that pluripotent stem cells had a single origin in the stem animal (Fig 1A) and that these cells are homologous across metazoans. Even if there were support for the homology of pluripotency across animals (which there is not, see: Srivastava et al., 2021 Ann Rev Cell Dev Biol) the OSKM genes that drive embryonic pluripotency in mammals are certainly not conserved across animals. The authors point to cnidarians and planarians as examples of other animal taxa that have clear pluripotent stem cells but both of these taxa lack orthologs of Oct4 and Nanog and the unusual pluripotent stem cells of Hydra seem to be specific to that lineage of cnidarians, rather than an ancestral trait of the phylum (see: Chapman et al., 2010 supplementary information). The authors erroneously point to several citations to suggest an implicit relationship between sox/pou genes in cnidarians and placozoans but none of the papers they cite actually demonstrate evidence of sox/pou co-expression (rather, they just demonstrate both gene families are present in these taxa). Furthermore, recent evidence from another cnidarian (Nematostella) suggests SoxC is the multipotent stem cell regulator upstream of the more restricted SoxB genes and there's no evidence pou genes are so expressed with SoxC or soxB genes in this system. Suggesting there is an (undemonstrated) ancient interaction between sox and pou factors is misleading to readers that may be unfamiliar with the biology of these early-diverging metazoan lineages.

In addition to the conceptual concerns outlined above, there is generally insufficient detail to evaluate much of the data presented in this manuscript. The authors provide the number of colonies that express GFP in their in vitro experiments rather than a percentage of total colonies, which makes these results hard to discriminate from noise. They also provide some measure of “efficiency” in a heat map but do not explain how this metric was calculated. The only evidence of actual pluripotency is demonstrated by immunofluorescence (Fig 2D,E) but these figures include neither positive controls (with mouse Sox2) nor negative controls (without primary antibody). For cooperativity assays, the authors fail to explain the axes in their figures (Fig 1F,G) and while they say binding was “highly specific” it’s not clear how this was determined. Again, it seems a different control would have been more informative, for example what would Cic look like in these same assays? The authors clearly have a lot of experience with cell culture but if this manuscript is marketed for a broad readership, the figures and methods will need be presented with much greater transparency.

Specific comments:
No citations are provided to support how the authors identified “key” residues in Sox and POU orthologs (Fig 1C and 5B)
Many of the methods are vague - What is a “round” of picking for data in Fig2 and S3? What constitutes an “experiment” when referring to replicates? What, specifically, do the authors consider to be “biological” vs “technical” replicates?
HMG and POU domains were validated with PFAM database but no indication of the significance cutoff (E-value) is

provided.

Fig 1C – do the logos indicated “unicellular holozoans” include M. brevicolis and S. rosetta or just the new taxa presented in this manuscript? It would be helpful if the authors could also include a logo for non-Sox HMG boxes for comparison.
Fig 3C,D - The binding assay figures look identical for both taxa, how does this indicate different cooperativity? How does this translate into different abilities to induce stemness? Also the figures need to be explained more fully – what are the numbers and +/- signs at top indicating?
Fig 3E – what would the cooperativity of mouse/human Cic look like? What about sox from a sponge or cnidarian or planarian?
Fig S1 – the previously identified sox genes from Mbrevicolis and Srosetta should be clearly labeled to match those sequences in the alignment in Fig S2. There are only two sox-like genes for each species in the alignment but there are three for Srosetta in the tree. (Same for Pigoraptor – sequences in tree should also be represented in the alignment.) Taxon abbreviations should be indicated. Some of the sequences are just indicated as “contig_XXX”.
Table S1 is not provided
Fig S7A – ancestral state posterior probability plots – these are not explained at all. What do the probabilities correspond to - specific residues? Whole HMG binding domains? The authors need to provide the actual sequences that were cloned and expressed in cell culture and explain how fluorescence was quantified (Fig 4C SoxCEF looks fluorescent to me).

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have extensively revised the manuscript, adding additional experiments and analyses, and removing many of the most (evolutionarily) speculative statements. This has resulted in a substantially improved manuscript.

I have no remaining major concerns, although I feel that the overall interpretation of the POU results, while improved and substantiated with the new POU3 experiment, is still biased towards trying to fit the vertebrate Sox-POU model into a distant ancestor. To do this thoroughly, given the lack of compelling resolution at critical nodes of the POU phylogenetic tree, the authors would need to do array of experiments with diverse ancestral POU family members. This is perhaps beyond the scope of this study. Instead, the authors should tone-down their extrapolations.

Reviewer #2

(Remarks to the Author)
The authors have fully adressed my questions and comments, and I am happy with the improved version of the manuscript.

Reviewer #3

(Remarks to the Author)
The authors refer to the biochemical properties of Sox genes to regulate iPSCs as "already present" in choanoflagellates but this is a misrepresentation of the data. They have not shown this property to be shared by SoxB genes from any of non-bilaterian taxa (which would support their implication of continuity) and they imply that the properties exhibited by extant choanoflagellates necessarily represent the properties of these proteins in the choano/animal ancestor. This language needs to be modified. Fig 1A still implies that stem cells are homologous and have a single origin in the common ancestor of animals, despite the claim that this is not implied in the text. The authors still use the generalized language of "stem cells" when describing non-bilaterians but do only a cursory job of explaining that these cells do not fit the pattern of an ancestral SoxB/POU relationship.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

General comments

This report describes the discovery of members of Sox and POU transcription factor families in choanoflagellate and filasterean unicellular holozoans. This pushes the origin of these families back beyond the origin of the metazoans, which has been suggested in previous publications. Members of both families – Sox2 (SoxB) and Oct4 (POU5) specifically - are involved in inducing stemness in vertebrates, and have been shown or implicated to be involved in stemness in other bilaterians and cnidarians.

The evidence for the ancestral holozoan Sox gene having biochemical properties remarkably similar to extant Sox2 is compelling and very interesting. Multiple lines of phylogenetic and experimental evidence support the authors' claim that genes identified in choanoflagellates can substitute for mammalian Sox2 in a range of molecular and cell contexts, including partnering with mammalian Oct4 in DNA-binding analyses using the canonical (mammalian) SoxOct DNA cis-element and inducing mouse PSCs. Given the similarity of the choanoflagellate Sox sequence to metazoan Sox2/SoxB and that these HMG DNA-binding domains were expressed as chimerics that were flanked by mammalian Sox2 C- and N-sequences, it might be expected that these Sox hybrids will be functional in mice.

Response: We wish to point out that we have not only analysed 'Chimeric' Sox hybrids but also four full-length choanoflagellate Sox factors from *Salpingoeca helianthica* and *Mylnosiga fluctuans* (Figure 2, S3C). Only a small set of initial experiments used the 'hybrid Sox' and once we discovered that full-length choanoflagellate Sox factors can also substitute for mouse Sox2 in iPSC generation, we used full-length factors for key experiments such as the generation of chimeric mice. Even with N- and C-terminal flanks that are not conserved at all, choanoflagellate Sox can replace mouse Sox2 and generate bona fide iPSCs.

In a set of similar analyses it was shown that the choanoflagellate POUs do not function as the mammalian Oct4 despite possessing bona fide POU and homeodomains with conserved residues responsible for DNA binding. The phylogenetic tree shown in Fig. 5A differs from previously published trees of the POU family that used only metazoan sequences, with sponge and fly sequences sitting outside the recognised 6 subfamilies, and cladeing within or outside the choanoflagellate POUs. It appears this topology may be "forced" by the parameters set by the authors and described briefly in the Methods

(there is no further analyses in the supplementary materials except an alignment of POU and homeodomain sequences, which is convincing in terms of the choanoflagellate sequences being genuine POU HDs but not it terms of their relationship to metazoan POUs). The phylogenetic analyses should be redone and the results reported in detail. Choanoflagellate POUs may not have worked in the mouse system because they are highly divergent.

Response: We apologize for the confusion caused by Figure 5A. This is a tree that includes all homeodomains in a subset of holozoans, and its main purpose was to show that a homeodomain-specific phylogeny places choanoflagellate POU sequences as sister group to animal POU sequences. The *Drosophila* sequence that appears in Figure 5A is not a POU member, but a ZFHX member (Zhfx1). As suggested by the reviewer we have made new focused trees, see responses below, where our topology of POU subfamily evolution is consistent with previous publications, recovering all the major groups (POU1, 2, 3 / 5, 6), and where choanoflagellate POU sequences do not cluster with any of the animal-specific subfamilies. New sponge and ctenophore data corroborate that the last metazoan ancestor possessed POU1 and 6 members, and depending on the position of sponges (or ctenophores) as a sister group to the rest of animals, then POU3 would also be part of that urmetazoan repertoire (new Figure S13).

Even though the authors convincing demonstrated that the ancestral choanozoan (choanoflagellates + animals) Sox gene was likely to have been a member of SoxB/Sox2 subfamily, their interpretation of its role in the evolution of stemness in confusing in places. As written, some readers might infer that the origin of stemness and Sox2 are being conflated. The narrative around the evolution of metazoan pluripotency and potential role of Sox needs to be explained more clearly (see specific comments below). For instance, both Sox and POU genes were detected in only a small number of unicellular holozoans, and it is inferred that these genes were lost in most choanoflagellate and filasterean lineages. If this is the case, is there any aspect about the life cycles of the species these genes were found that might provide insights into role of Sox in these species and thus the role of the ancestral Sox and its potential to be involved in cellular reprogramming or state changes. Related to this point, a discussion on the fact that most of the unicellular holozoan species are represented by a transcriptome(s) and not a genome is warranted. Are these transcriptomes targeted stages of the life cycle or opportunistic (random) sampling? Also, does depth and coverage of sequencing contribute to the inability to find Sox and POU genes in most species?

Response: We have included extra information on the quality of the choanoflagellate transcriptomes, see below, and we have also included the potential caveat of missing

genes being the consequence of having incomplete transcriptomes or genomes in the discussion.

Specific comments

Abstract

1. Sox genes shows that their molecular ability to induce stemness was already present in the last common ancestor of animals and their unicellular relatives.

This inference does not necessarily follow. Just because an ancestral Sox gene can function in contemporary mammals does not necessary mean that was its ancestral function.

Response: The reviewer is right that this sentence can be confusing. We have rephrased it as
"Ancestrally reconstructed Sox proteins indicate that iPSC formation capacity is pervasive among resurrected sequences … "

2. Our findings imply that the evolution of stem cells exploited a pre-existing set of transcription factors, where the critical innovation involved an initial change in DNA specificity of POU and the exaptation of the ancestral capacity to interact with Sox transcription factors.

Confusing as written. Is "exploited" the best term to describe this situation? Also, the statement infers that the partnering of SoxB and POU5 was instrumental in the early evolution of stem cells. There is no evidence in this report of these two TFs partnering in unicellular holozoans. Nor is there currently evidence of them partnering in sponges, placozoans, ctenophores or cnidarians, despite the presence of stem cells, SoxB and multiple POUs in these taxa. As such, this statement needs to be tempered and rewritten.

Response: We have avoided the use of "exploited", we now use exaptation. The reviewer is right that at this stage we cannot assume that SoxB and POU5 interactions occur in any early animal phyla. We do not imply that POU5 (Oct4) was ancestral, as it is clearly a vertebrate innovation. POU5 arose from a gene duplication from a more ancient POU3 member and POU3 factors are found all the way to sponges and cnidarians. There is published evidence of POU3 members interacting with SoxB in *Drosophila* (PMID: 10844029 and 24314314) and in mice (PMID: 26265007 and 23437007), thus the POU3/SoxB partnership could be ancestral to animals. To examine

whether unicellular Sox could interact with mammalian POU3, we have performed an additional series of experiments and showed that Salhel Sox-I dimerises with mouse Brn2 the same way mouse Sox2 does (new Figure 3G). This leaves open the possibility that partnering of SoxB with POU3 was important in the early evolution of stem cells. We have rephrased this whole ending like:
“Our findings imply that the evolution of animal stem cells involved the exaptation of a pre-existing set of transcription factors, where pre-animal Sox was already biochemically similar to extant Sox, whilst POU factors required evolutionary innovations.”

Introduction

1. We propose that the Sox/POU partnership emerged alongside the trajectory towards animals and was likely driven via molecular exaptation of the previously established Sox-POU-DNA dimerization and binding capacities.

See comments above - Abstract 2

Response:
We have added additional data showing that choanoflagellate Sox forms a dimer with mouse POU3 with similar efficiency as mouse Sox2 (new panels in Figure 3). This raised the possibility of a SoxB/POU3 at the onset of animal evolution, as POU3 is found in sponges. Almost all animal POU factors have a cysteine at position 51 of the POU homeodomain whereas choanoflagellate sequences have a glutamine. This is likely one of the changes that allowed animal POU factors to bind octameric ATGCAAAT sequences which is a pre-requisite for the DNA-dependent cooperativity with Sox factors to regulate genes in multipotent and pluripotent stem cells in mammals. In light of our new data showing a possible interaction of choanoflagellate Sox with POU3 factors, we wish to still argue in favour of our proposed model that SoxB/POU3 interaction might have evolved early in animal evolution, yet we also acknowledge that other scenarios and further experiments will be needed to corroborate this model.

Discussion

1. Their molecular features to bind and open up closed chromatin as pioneer factors might have primed Sox proteins for becoming stem cell regulators in animals 4,5,7,31.

Although this is an interesting deduction based on modern Sox-chromatin interactions that seems very plausible, cellular plasticity – allowing cellular responses to exogenous and endogenous signals – likely evolved before a cell state that we recognise as a pluripotent stem cell. Consider reworking this sentence based on these and comments above (Abstract 2).

Response: The reviewer is right pointing that cellular plasticity predates multicellularity. The point we want to raise here is that probably these molecular features of Sox are what made them successful cell fate programming factors in animals: their capacity to reach “closed” heterochromatin parts of the genome to activate silenced genes. Other TFs would be less good at that as the work of Ken Zaret and others has demonstrated. Even if stem cells are not directly homologous across animal lineages, finding SoxB members in many stem and progenitor cell lineages across many animal clades implies that Sox factors might have either an ancestral and highly conserved role in these cells - or they have been convergently deployed for the same functions more than once. We have rephrased the sentence to make it more nuanced, and contemplate both possibilities in the discussion.

2. Our data show that many of the biochemical features that render Sox stemness factors are already present in the last common ancestor of animals and choanoflagellates, facilitating its co-option into pluripotency networks (Figures 1E-G,3A-F).

The first part of the sentence is correct but the second half does not follow. As written a pluripotency network was in place without the need for Sox. There seems to contradict other statements in the manuscript that implicate Sox being central to the evolution of pluripotency.

Response: We have rephrased the sentence as it was prone to misinterpretation. We have also changed pluripotency for stem cell regulation, as not all SoxB are necessarily marking pluripotent stem cells, but other tissue specific stem cells.

3. As a notable exception, sponges from the demosponge class do not express SoxB in the assumed pluripotent stem cell in this lineage, the archeocytes 55,56, yet this might not extend to other sponge classes.

SoxB is expressed in larval and adult archaeocytes in Amphimedon – see supplementary figures and tables in References 1 and 56.

Response: We thank the reviewer for pointing this out and have rephrased this sentence. We were led to confusion due to Figure 5d in reference 56 (numbering changed in revised manuscript; PMID 29942020), as SoxB is a stronger marker for pinacocytes, but it is indeed also expressed in archeocytes.

4. Thus, given the capacity of a choanoflagellate Sox to interact with Oct4 to drive pluripotency, analogous partnerships might have enabled animal evolution (Figure

Not sure what this means - need to be more explicit. New TF partnerships driving multicellular complexity? For example, bZIPs?

Response: We refer to the partnerships between SoxB and POU3 factors. Following the reviewer's suggestion, we rephrased this sentence to make more explicit what we mean.
"Thus, given the capacity of a choanoflagellate Sox to interact with the POU3 factor Brn2 (Figure 3e,g), analogous partnerships between Sox and octamer binding POU members might have been important for the evolution of stem cell regulatory networks (Figure 6b)."

5. Since Sox pre-dates the origin of stem cells, the remaining question is if there was any further molecular innovation required for the establishment of this gene regulatory network in animals.

Again, vague in terms what is meant evolutionary. This seems to be a difficult question to answer given the paucity of GRN information from early-branching animal lineages

Response: We have also changed the sentence to avoid confusing interpretations. Now it reads: "*Since Sox pre-dates the origin of stem cells, the remaining question is when a Sox driven stem cell gene regulatory network evolved in animals.*"

Methods

1. We used the human Sox2 protein as query for a BLASTP search against the predicted proteomes of unicellular holozoans, including 22 choanoflagellates (20 transcriptomes and 2 genomes), 4 filastereans, 7 ichthyosporeans …

The lack of Sox and POU genes could have to do with transrcriptome coverage. It would good to see the sequencing data. Are the transcriptomes stage specific?

Response: The reviewer raises a fair point, as a transcriptome is not, in principle, as complete as a genome. We just want to highlight that the choanoflagellates we have profiled in this manuscript were generated in previous studies by other authors (see PMID: 29848444 and PMID: 31624206) and are publicly available. These choanoflagellates might have more than one stage present in culture (colonies, rosettes, stalked cells, etc), while others might not fully represent the full transcriptional potential of the species, which is an unavoidable issue when working with protists. However, we have run BUSCO completeness analysis on all choanoflagellate datasets, and whereas transcriptomes all have >87% BUSCO completeness, the genomes of *Monosiga brevicollis* and *Salpingoeca rosetta* have lower BUSCO scores (78% and 83% respectively). Therefore, absence in these datasets cannot be ruled out as a technical or biological artifact, but both *Salpingoeca helianthica* (91%) and *Mylnosiga fluctuans* (93%) were done in comparable conditions to the other species and yet they express these genes.

We now include a supplementary table with the BUSCO results for each choanoflagellate dataset in the study:

01_Diaphanoeca_grandis.proteins.fasta C:92.2%[S:90.6%,D:1.6%],F:3.1%,M:4.7%,n:255
02_Acanthoeca_spectabilis.proteins.fasta C:93.3%[S:90.6%,D:2.7%],F:1.6%,M:5.1%,n:255
03_Salpingoeca_punica.proteins.fasta C:90.2%[S:89.8%,D:0.4%],F:6.3%,M:3.5%,n:255
04_Salpingoeca_urceolata.proteins.fasta C:92.6%[S:87.1%,D:5.5%],F:4.3%,M:3.1%,n:255
05_Hartaetosiga_gracilis.proteins.fasta C:88.3%[S:86.3%,D:2.0%],F:5.9%,M:5.8%,n:255
06_Salpingoeca_macrocollata.proteins.fasta C:89.0%[S:85.9%,D:3.1%],F:5.9%,M:5.1%,n:255
07_Stephanoeca_diplocostata.Australia.proteins.fasta C:88.2%[S:84.3%,D:3.9%],F:2.0%,M:9.8%,n:255
07_Stephanoeca_diplocostata.France.proteins.fasta C:90.2%[S:88.2%,D:2.0%],F:3.1%,M:6.7%,n:255
08_Helgoeca_nana.proteins.fasta C:93.7%[S:91.0%,D:2.7%],F:1.2%,M:5.1%,n:255
09_Salpingoeca_kvevrii.proteins.fasta C:94.9%[S:87.1%,D:7.8%],F:3.1%,M:2.0%,n:255
10_Didymoeca_costata.proteins.fasta C:94.9%[S:92.5%,D:2.4%],F:2.0%,M:3.1%,n:255
11_Choanoeca_perplexa.proteins.fasta C:92.1%[S:89.4%,D:2.7%],F:3.1%,M:4.8%,n:255
12_Salpingoeca_infusionum.proteins.fasta C:91.0%[S:87.1%,D:3.9%],F:3.9%,M:5.1%,n:255
13_Microstomoeca_roanoka.proteins.fasta C:93.3%[S:90.6%,D:2.7%],F:2.7%,M:4.0%,n:255
14_Savillea_parva.proteins.fasta C:87.5%[S:85.5%,D:2.0%],F:6.3%,M:6.2%,n:255
15_Hartaetosiga_balthica.proteins.fasta C:90.2%[S:89.8%,D:0.4%],F:4.7%,M:5.1%,n:255
16_Salpingoeca_dolichothecata.proteins.fasta C:90.6%[S:88.2%,D:2.4%],F:5.9%,M:3.5%,n:255
17_Codosiga_hollandica.proteins.fasta C:88.7%[S:76.9%,D:11.8%],F:5.9%,M:5.4%,n:255
18_Salpingoeca_helianthica.proteins.fasta C:91.4%[S:89.8%,D:1.6%],F:4.3%,M:4.3%,n:255
19_Mylnosiga_fluctuans.proteins.fasta C:93.3%[S:90.2%,D:3.1%],F:3.9%,M:2.8%,n:255
Choanoeca_flexa.fasta C:89.8%[S:82.0%,D:7.8%],F:3.9%,M:6.3%,n:255
Monosiga_brevicollis.fasta C:78.8%[S:78.8%,D:0.0%],F:7.5%,M:13.7%,n:255
Salpingoeca_rosetta.fasta C:83.1%[S:83.1%,D:0.0%],F:6.7%,M:10.2%,n:255

(C: complete, [S: single copy, D: duplicated], F: fragmented, M: Missing)

We also discuss the possibility that future genome sequencing efforts might reveal more Sox or POU sequences in choanoflagellates (or other holozoans).

2. The resulting 756 sequences were aligned using MAFFT, trimmed uing trimAl, and IQTREE to build the phylogeny…

Why isn't this shown? Using POU6 as an outgroup appears to force sponge and othermetazoan POUs outside the recognised metazoan subfamilies. It is unclear why this was done.

Response: We apologize for the lack of detail in this section. The reviewer got confused by the figure 5A, where we did not use POU6 as an outgroup, but we performed a phylogenetic tree of all homeoboxes present in unicellular holozoans as well as in few animal species (*Homo sapiens, Drosophila melanogaster, Nematostella vectensis, Amphimedon queenslandica, Trichoplax adhaerens, Mnemiopsis leidyi*). The purpose of that tree was to test if the homeodomain found in choanoflagellate POU sequences would cluster with animal POU factors when only using Homeodomain amino acid positions, discarding an unlikely but potential convergent evolution of POU-specific and POU-homeodomains in choanoflagellates and animals. Given that the phylogeny of homeodomains is really complex, the topology of the tree including all sequences from all major types (TALE, ANTP, PRD, Six, etc) is unlikely to be as accurate as a more focused phylogenies on any given family. This is why we have now included another phylogeny including more animal POU sequences, with and without several outgroups (Six LHX and ZFHX homeoboxes, and an unrooted tree), where we use both POU domains (POU- specific and POU-homeodomains) for constructing the phylogeny, therefore having more positions than in the tree in Figure 5A (new Figure S13). These focused phylogeny topologies recapitulate previous publications on POU evolution, however, the sequences from choanoflagellates, and few sequences from two demosponges (*Amphimedon queenslandica* and *Tethya wilhelma*) have very long branches, which place them in an indeterminate position among established POU families. Therefore, the new trees now found in Supplementary Figure S13 supports choanoflagellate sequences belonging to POU, and they cannot be classified as any particular animal subfamily. If the branching with the sponge orphan POU factors is biologically meaningful or a long branch attraction artifact would probably require more sequences from sponges and unicellular holozoans. Of note, neither full genomes of calcarean, hexactinellid or homoscleromorpha sponges possess copies of these “Orphan” POU, this suggests these are divergent orthologues only found in few demosponges.

3. Protein samples and fluorescently labelled DNA were mixed in EMSA buffer…

Rephrase to make clear these are recombinant proteins, not protein samples from cells.

Response: We have included “Purified recombinant protein samples” to avoid potential confusion.

Reviewer #2 (Remarks to the Author):

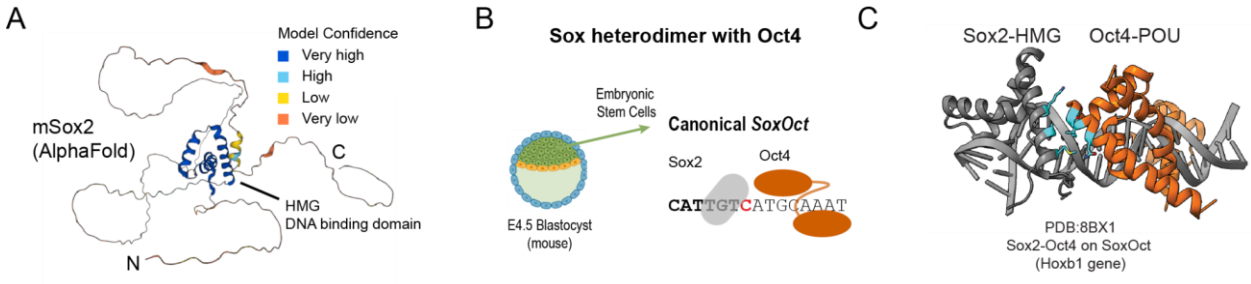
The manuscript reports the striking ability of pioneer transcription factors Sox orthologues from unicellular organisms like choanoflagellate and filasterea, to interact with the same DNA motif like the mouse Sox2, and even the ability to induce iPSC formation. This is an exciting finding, and of great interest to a scientific communities interested in chromatin, transcription factors, reprograming, and evolution. The findings shed light on the early evolution of pioneer transcription factors, and pluripotency. There are a couple minor comments that I would like the authors to take care of.

- A question related to the reprogramming experiments with Sox orthologues. I think it would be really important to show how the levels of Endogenous SOX2 are not changing in different reprogramming conditions - when different Sox orthologues are used in the Yamanaka cocktail. A control with a temporary mouse Sox2 knockdown and the addition of Yamanaka cocktail should be performed (at least for one Sox2 orthologue). In that case reprogramming should be still observed. Or any alternative experiment to show that the level of endogenous Sox2 is close to zero, and does not change, so that we are sure that the observed differences are clearly attributed to the introduced Sox and not the endogenous.

Response: To address this valid point, we conducted RT-qPCR analysis using primers specific for the endogenous Sox2 mRNA at three different time points during reprogramming (day 6, 9, and 12) with the full-length *Salhel* and *Myflu* Sox variants as well as in established clonal iPSC lines and the starting MEFs. None of the four Choanoflagellate Sox factors show elevation in endogenous Sox2 expression until day 9 (Supplementary Figure 4A). At day 12, a time when first GFP-positive colonies have appeared, endogenous Sox2 gradually becomes detectable albeit at a very low level. We only detected high levels of endogenous Sox2 expression in established iPSC lines. These findings demonstrate that the pluripotency induction with unicellular Sox factors proceeds without an initial interference from endogenous Sox2. Then, once pluripotency is induced, exogenous ‘Viral’ Sox factors are silenced and endogenous Sox2 maintains the pluripotent state.

- Another important control would be to add a chimera with mouse Sox2 HMG and Salhel and/or Phi N and C-termini, and see if that chimera can lead to iPSC (as in Fig. 2A), in order to complete a picture of the role of DBD vs tails in the pluripotency induction. And potentially a control with a Sox2 DBD only (with tails cut off).
Response: To perform these interesting experiments, we constructed three different variants of Sox2: Sox2-HMG only (without N- and C-terminal tails or flanks), Sox2-HMG flanked by the Salhel-Sox-I N- and C-terminus (Salhel-I-Sox2), and Sox2-HMG flanked by the Salhel-Sox-II N- and C-terminus (Salhel-II-Sox2). The variant consisting solely of the DNA binding Sox2-HMG domain only was unable to induce pluripotency. For the chimeric versions of Sox2-HMG with the Salhel-Sox-I and Salhel-Sox-II regions, we successfully observed the generation of iPSCs, but the efficiency of iPSC generation was reduced compared to that of the wild-type mSox2 (new Supplementary Figures 4B, C). We conclude that the DBD of Sox factors determines their capability of pluripotency induction, while the N and C-termini contribute to the overall efficiency, which also aligns with our previous publication (PMID: 37607832). It is necessary to have flanking tails to support iPSC generation as their complete removal is detrimental. Yet, diverse and poorly conserved N and C-terminal tails can support the functions required for successful reprogramming.

- Could the authors also discuss a bit, which part of Sox is the most essential from the evolutionary stand point in terms of pluripotency induction in their opinion, and why? Is it the DBD that interacts with Oct4, or are those the unstructured tails? Based on mutagenesis, one would think it's the DBD, but the chimera of Salhel DBD + native Sox2 tails could not lead to iPSC. Please comment on that.



Reviewer Figure 1. (A) AlphaFold2 predicted structure of full-length mSox2 with model confidence color-coded. (B) Schematic of Sox2/Oct4 heterodimer formed on the canonical SoxOct DNA element found in pluripotency enhancers from pluripotent cells such as embryonic stem cells at the inner cell mass of a blastocyst. (C) Structure of the ternary complex of Sox2/Oct4 DNA binding domain heterodimer on the canonical SoxOct DNA of the *Hoxb1* gene locus (PDB ID: 8BX1).

Response: We may have caused this confusion here:, The Salhel DBD + native Sox2 tails can reproducibly generate iPSCs (Figure 3). For Sox proteins, the N/C termini are unstructured and poorly conserved (Rev Fig. 1A) while HMG DNA binding domain (DBD) is structured and highly conserved (see Figure S3C for domain-wise conservation). The HMG domain of Sox interacts with the POU DBD of Oct4 - there is no biochemical or structural evidence that the unstructured tails contribute to this interaction (Rev Fig. 1B-C). This direct interaction of the Sox and POU DBDs is critical for these proteins to form DNA-dependent heterodimers on the canonical SoxOct DNA element found near many pluripotency genes. Yet, Sox factors still require unstructured transactivation domains (TADs) but as we showed here and in previous studies, TADs with highly diverse sequences can support reprogramming albeit with different efficiency (see new Figure S4B,C see also (PMID: 37607832)). We have revised the manuscript for clarity and added the reciprocal domain experiments the reviewer suggest below.

- Please show the complete DBD sequence of the mouseSox2, Sox17 and the orthologues in the figure1 (instead of S1), and indicate sequence identity scores just for the DBD. The partial (cut out) sequence is misleading. Also please show the complete alignment including the unstructured regions in the supplementary figure 1. Please comment whether there is any sequence similarity to mouseSox2 in the unstructured parts.

Response: Thank you for this comment, we understand the preference for having the alignment of the DBD in Figure 1 instead of S2. We have tried to integrate the complete sequences but to create the logo we had to remove gaps in alignment and found that showing the key sequences best serves the purposes of quick comprehension in the main figure. As an additional reference, we have added TCF/LEF family members to Figure 1C. The supplements now contain the colour-coded alignment of the full DBDs which includes an expanded set of sequences (Figure S2A). The per residuec alignment scores are also shown in supplementary figures S2A. The unstructured areas outside of the DBD have very poor conservation with many gaps that are not very informative. Such an alignment would fill several pages. We have uploaded this alignment to the figshare repository.

- Please comprehensively color code Figure S1 seq alignment, there is no back-and-white requirement.

Response: Thank you for this suggestion, we have colour-coded all alignments in the manuscript according to the Clustal X colour-scheme (https://www.jalview.org/help/html/colourSchemes/clustal.html) .

- Please also discuss whether the sequence similarity between mouseSox2 and different Sox orthologues is in agreement with how “they act”: i.e. a higher seq similarity in those orthologues that also induce pluripotency and vice versa.
Response: In terms of orthologs, mouse Sox2 can be replaced with human Sox2. The Hitoshi Niwa lab has shown that many distant invertebrate SoxB orthologues can substitute for mouse Sox2 in mouse ESCs (PMID 27582319). In contrast, among mammalian SoxB paralogs, only SoxB1 group members (Sox1,2,3) can support pluripotency reprogramming but other paralogs (including SoxC,D,E, F) cannot (PMIDs: 18059259, 23963638). This highlights how just looking at sequence similarity is not enough to understand how a gene might “act”. This said, SalHel and MylFlu sequences have higher similarities to Sox2 than the SalRos and MonBre Sox-like sequences, with the exception of Mylflu-Sox-II that is 42.5% where Salro Sox-like is 42.9%, yet this does not make it good enough to substitute mSox2, while Myflu-Sox-II is capable.

- Fig. 5B need a complete sequence of the domain, and not a cut out; same in Fig. 4B. Or remover them completely if the sequences are shown later in the supplement.

Response: Thank you for this comment, we understand the preference for having the alignment of the POU domain in Figure 5 instead of the supplements. However, this may fill a full page and we could not find a way to integrate such a panel. We therefore wish to retain subset for clarity and readability. To address the comments, we have improved our alignment in Figure S12 (was S9 previously) with additional annotations, the suggested colour coding and additional protein sequences. We are open to move Figure S12 to the main text but prefer keeping it in the supplement to avoid overcrowding the manuscript. We wish to include the truncated logos rather than removing as it helps us to illustrate features of unicellular POU (especially within the POU_S) that are also conserved animals and those that differ (such as residue 50 of the POU_{HD}).

- Typo on page 4: Holozoa (instead of Holoza)

Response: We have corrected the typo in the manuscript.

Reviewer #3 (Remarks to the Author):

Summary
This study investigates the evolutionary origins of pluripotency by examining the functions of two pluripotency genes (Sox2 and Oct4) from organisms outside metazoa. The authors demonstrate the presence of Sox-family HMG box genes and POU family

genes in choanoflagellates and show that the Sox genes can bind DNA and activate the expression of Oct4 in mouse embryonic fibroblast cultures. They use this result to suggest that the pluripotency-inducing capacity of Sox genes evolved before the last common ancestor of animals. The approach is bold and creative but I am not convinced that the data they present tell us anything about OSKM-driven pluripotency, which is a vertebrate specific phenomenon. Several concerns are outlined below.

Response: We are glad the reviewer appreciates our 'bold and creative' approach. We wish to politely point out that unicellular Sox do not only activate Oct4 expression in some reporter assay. Rather, they can substitute for Sox2 to convert somatic cells into induced pluripotent stem cells (iPSCs) with all molecular and functional hallmarks of pluripotency and self-renewal typical for blastocysts derived embryonic stem cells including their ability to contribute to life chimeric mice (please see Reviewer Figure 2 below and further explanations). It was initially shown by Shinya Yamanaka (who discovered the iPSC technology in 2006 and received the Nobel Prize in 2012) that Sox2 is essential for reprogramming and cannot be easily substituted for by paralogous Sox factors (except for fellow SoxB factors Sox1 and Sox3 PMID: 18059259). We later on confirmed these findings by examining additional Sox factors and mutants thereof (PMID: 23963638). Hence, it is by no means trivial that mouse Sox2 can be substituted for by a unicellular Sox to achieve the 'rolling back' of embryonic development associated with the captivating conversion of a specialized somatic fibroblast cell into a pluripotent stem cell with properties of the stem cells otherwise only found in the inner cells mass of blastocyst stage embryos.

General comments

The main conclusion is that bona fide sox genes were recovered from the two new choanoflagellates (Salpingoeca helianthica and Mylnosiga fluctuans) where only sox-like HMG box genes have been discovered previously (in the two previously sequenced choanoflagellates S. rosetta and Monosiga brevicolis). The authors support this claim using phylogenetics to group the new choanoflagellate sox genes with animal sox genes (to the exclusion of the sox-like genes from S. rosetta and M. brevicolis) and by demonstrating that Salhel sox can replace mouseSox2 in assays of induced pluripotency. From the tree provided (Fig S1) the rationale for including the Salhel and Myflu sequences in the Sox family while relegating the Sros and Mbre sequences to "Sox-like" is not clear. There is very low bootstrap support (~30%) connecting Salhel and Myflu to the animal sox genes and there appear to be some sponge sequences (Scil) that were excluded from the sox clade. Examining the partial alignment provided in Fig S2, it also does not seem there is a clear reason for excluding the sequence labeled Srosetta-Sox-II from the functional (in vitro) analyses, as it shares all the same "key" residues as the sox genes from human Sox2, Salhel Sox, and MfluSox. If the

authors are going to claim they have found something unique to Salhel and Myflu, they would need to perform the same experiments (in mouse embryonic fibroblasts) on (minimally) the Sros-SoxII sequence and demonstrate that it lacks this capacity.

Response: To address this comment, we accordingly constructed Sox variants derived from *Salpingoeca rosetta* (Sros) and *Monosiga brevicollis* (Mbre) to investigate their binding to the Sox motif and their potential for pluripotency induction. Our results show that Sros Sox exhibited only very weak binding to the Sox motif compared to Salhel or Pchi (~30-fold lower affinity, Figure 1H, Supplementary Figure 2E, F). Mbre did not show any binding to the Sox motif, even at high protein concentrations. In our reprogramming tests, we observed that Mbre was unable to induce pluripotency in any of the 9 replicates. As for Sros, we only observed 1 GFP-positive colony in 9 replicates. However, the colony could not be maintained after picking, indicating that it was not fully reprogrammed and does not represent a self-renewing iPSC colony (Supplementary Figure 3). We thank the reviewer for suggesting these indeed very interesting experiments which helped us to experimentally establish that Salhel and Myflu Sox-like differ from bona fide Sox factors in biochemical and functional assays.

Additionally, they need to provide better support for the topology they have chosen to represent with their Sox tree by performing an approximately unbiased test (see: Shimodera 2002 Syst Biol) and providing intermediate trees that were generated before they decided to prune sequences.

Response: We acknowledge the suggestion to describe in more detail how the maximum likelihood (ML) phylogenetic tree, shown in Figure S1, was inferred. In the revised manuscript, we added the information that we repeated the tree search ten times using the same input parameters (but different seeds). The tree shown in Figure S1 had the highest log likelihood, which was almost identical to the second and third-best trees (-14520.932, -14520.933 ,-14520.935). These trees also did not differ in their topology but only showed minor differences in their branch lengths. We added a description of this procedure to the methods section. The finding that the holozoan Sox proteins of Salhel, Myflu, Pvie, and Pchi form a sister clade to the metazoan Sox proteins encouraged us to classify these hits as Sox proteins. Nevertheless, we are aware that the inferred topology of the Sox/Sox-like clade deviates from the holozoan species tree, which would imply either incomplete lineage sorting or a set of additional gene duplications and losses. We, therefore, also performed a set of constrained tree (CT) searches ensuring that all Sox/Sox-like sequences followed the holozoan species tree topology (that is Animals,Choanos,Filastereans) and - following the reviewer's request - tested these trees against the ML tree with an approximately unbiased (AU)-test. The AU-test did not reject any of the tested CT trees, implying that the applied

constraints are not ruled out by the data. We, therefore, also repeated the ASR on all 5 tested CT trees and found that the signature residues that mediate the base-readout of the Sox DNA element are the same in the inferred ML and CT ancestors. We now also discuss the possibility of Sox-like being divergent Sox that have accumulated so many mutations that phylogenetically they are placed outside of the main Sox clade. Since we have now demonstrated that Sox-like do not act like bona fide Soxes, either interpretation does not change the main conclusion of our manuscript and highlight the importance of the *M. fluctans* and *S. helianthica* sequences.

Independent of whether the HMG box genes they identify from Salhel and Myflu are sox or sox-like genes, it's not clear what we really learn from the in vitro experiments presented here. The authors show that some small number of mouse embryonic fibroblasts express Oct4 when they are transfected with choanoflagellate sox in place of mouse sox2 but there's no way for the reader to evaluate the significance of these results. As a negative control the authors perform the same experiment with Sox17 which is known to be incapable of inducing pluripotency in this system. Considering Sox17 is also expressed in the early embryo (at/near the same time as Sox2) and is responsible for driving differentiation of the embryonic endoderm, it is reasonable to assume there has been strong diversifying selection to keep the functions of Sox2 and Sox17 separate in mouse. Thus, Sox17 is not really the appropriate negative control. Rather, I imagine that the authors might recover the same low number of Oct4 expressing cells (as they did with Salhel and Myflu sox sequences) if they repeated these experiments with just about any soxB gene from any invertebrate in place of mouse Sox2. Indeed, their ancestral state analysis (Fig 4) suggests this should be true. (Although, see below for additional concerns about the ancestral state analysis.) The proper control then would be to mis-express the complete (not chimeric) sox ortholog from something like a planarian or a hydra in place of mouse Sox2. If the "pluripotency-inducing" capacity of these sox genes is equivalent to that of the choanoflagellate sox gene, what have we really learned? Neither planarians nor cnidarians use the OSKM gene cassette in their pluripotent stem cells (indeed, neither of these taxa even have a proper ortholog of Oct4 or Nanog) so independent of the behavior of their sox genes, it is not reasonable to imply that the capacity for mouse-like pluripotency is an ancient property of sox genes. What the authors have demonstrated is that the capacity of Sox genes to bind their target DNA sequences is highly conserved across animals, which is very cool, but extending the conclusion to say something about the origin of OSKM-pluripotency is unwarranted.

Response: We are delighted that the reviewer shares our excitement and thinks that it is 'very cool' that the biochemical properties of Sox genes are conserved beyond animals with regards to binding the cognate Sox element. We wish to stress that we

went beyond a mere biochemical evaluation and show using a functional assay that the unicellular Sox factors can also recapitulate SoxB-like activities and induce pluripotency in mice. This shows that unicellular Sox can navigate the chromatin of mouse cells and team up with mouse factors to activate the pluripotency network. We apologize for having caused some confusion for our initial failure to explain the iPSC reprogramming methods and associated assays for a broad readership. The reviewer seems to imply that we suggest that OKSM is ancestral to animals when this is not our interpretation of the data or it is not suggested anywhere in the manuscript. We acknowledge that Oct4 is a vertebrate-specific paralogue of the POU family. We speculate (using those words) that since POU3 (the paralogue of POU5/Oct4) can also interact with SoxB members, and even insects show Sox-POU heterodimers, this type of POU-SOX interactions might have played a role in early stem cells of animals. As a hypothesis, it cannot be disregarded yet, while we agree that a canonic OKSM or Nanog-dependent pluripotency networks are restricted to vertebrates. But for example, there is evidence of co-expression of SoxB and POU6 members in *Amphimedon queenslandica* archaeocytes (see reference PMID 29942020, Sebé-Pedros el al Nature Ecology and Evolution), yet these potential cooperative binding events would need further testing and are beyond the scope of this study. When SoxB and POU factors were co-opted into stem cell regulation remains debatable, but we now demonstrate that these TFs are older than previously thought. We have rephrased many sections of the manuscript to avoid this confusing interpretation regarding OKSM being ancestral. We also added additional data showing the cooperate dimerization of Salhel Sox-I with mouse Brn2 (a POU3 member) to clarify that this interaction is not restricted to vertebrate-specific POU family members but can also be seen for the ancient POU3 family (new Figure 3G).
We would also like to highlight that we used full (not just chimeric) Sox sequences for most of our iPSC reprogramming experiments. Then, finding that other invertebrate SoxB sequences could induce pluripotency in mouse stem cells would not be against our main claims, on the contrary, since SoxB is an animal Sox family, thus it is the expected equivalent behavior across orthologues. In fact, a pluripotency maintenance experiment has been done with the expected results using amphioxus and fly SoxB members (see PMID 27582319). In contrast, and more surprisingly, choanoflagellate Soxes are co-orthologues of the whole Sox clade, they can substitute Sox2 in iPSC reprogramming. This is not self-evident, as they could be more similar to SoxC or F, which are unable to perform this function (PMID: 18059259, 23963638) and are equally descendants of the original animal Sox. A good example is the POU example, where we show that the gene is there in choanoflagellates, but it cannot substitute Oct4 as it has different (and likely ancient) biochemical properties.

The low-level activation of Oct4 certainly does not derive from an ancient relationship between sox and Oct4 but rather seems like it might be noise (there is even low-level

GFP expression in the SoxCEF experiment). The authors would need to examine the capacity of choanoflagellate sox to interact with other pou genes to suggest there was an ancient relationship between these factors.

Response: We do not imply that Sox and Oct4 have an ancient relationship, we show that Sox from choanoflagellates can interact with Oct4 and we have also now shown that they can interact with POU3 (Figure 3G). We further show that ancestrally reconstructed Sox sequences from many early nodes in Sox evolution can induce pluripotency in mice. As many groups have demonstrated that Sox/Oct4 dimers are essential to induce and maintain pluripotency, we therefore propose that the ability to do so is an ancestral biochemical feature of Sox proteins (interaction with Oct4/POU3 and probably other POU members). Most likely for the substitution of the ancestral residue at position 57 with a glutamate, which abolishes iPSC generation in the context of Salhel-Sox (Figure S7B) and mouse Sox2 (PMID: 21472822), the iPSC reprogramming ability was lost in the lineage that led to SoxC/E/F families. Yet we disagree with the suggestion that our experiments reflect only “noise”, as we are not just looking at “Oct4” activation, but fully characterizing these stem cells to the point of making a chimeric mouse, please see response and details of our protocol below.

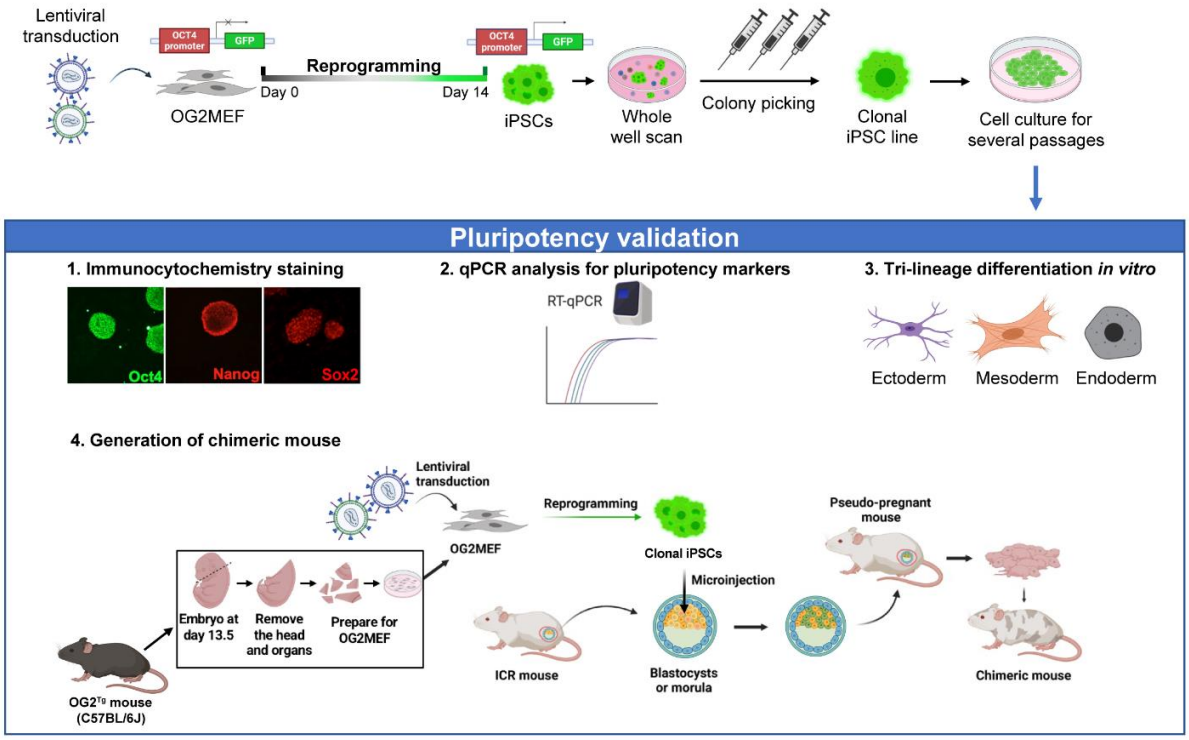
The title of the manuscript implies the authors have identified an ortholog of Oct4 in choanoflagellates, as other POU genes do not induce pluripotency (see: Nakagawa et al., 2008 Nat Biotech; Jin et al., 2016 Scientific Reports) but this is misleading. Oct4 is a POU5 ortholog and this clade of POU genes seem to be a vertebrate (or perhaps a deuterostome) innovation (Sukparangsi et al., 2022 Nat Comms; Gold et al., 2014 MBE).

Response: We agree that the title could be misleading, yet we did not imply that choanoflagellate POU is Oct4, and we acknowledge in the introduction and discussion that Oct4 is a vertebrate-specific paralogue. SoxB or even POU members have been identified as markers of stem cells, not necessarily pluripotent, such as neural or peptidergic stem cells in non-bilaterians, hence our use of “stem-ness associated” in the title. We chose those words to highlight why these TF families are important and relevant for a large community, and indeed motivated our analysis. Still, since it is true that not all Sox or POUs are associated with stem cells, we have removed that part from the title.

Importantly, the manuscript lacks a clear definition of how pluripotency is assayed in this system. In some places the authors imply that activation of Oct4-GFP is sufficient to conclude the pluripotent state has been reached, in other places they seem to suggest the expression of other factors is required (e.g., expression of Nanog and Esrrb are

shown in Fig2 and S5 but are never introduced or explained anywhere in the text of the manuscript).

Response: Thank you for your comments. We apologize for the confusion and have added the details accordingly in the manuscript. In response, we have also included a flowchart to illustrate the assays for pluripotency induction and validation (Reviewer Figure 2 and updated panel in Figure 2A) and enhance the readers' understanding of the experimental procedure in the revised text. The process of reprogramming MEFs into iPSCs is a result of a complex interplay of multiple factors. The overexpression of OSKM factors cooperatively modulates alterations in the epigenome and transcriptome, ultimately leading to the loss of the original cellular identity of terminally differentiated MEFs and their conversion into undifferentiated iPSCs with stem cell characteristics reminiscent to embryonic stem cells derived from inner cell mass of the blastocyst. While we utilize the OG2MEF reporter cell line, where the activation of GFP expression coincides with the endogenous expression of Oct4, it is important to note that Oct4 is only a first and indeed preliminary indicator of pluripotency. As a next step, it is crucial to assess whether the generated colonies can be established and maintained without the exogenous expression of OSKM after reprogramming. This assessment, often referred to as pluripotency validation in the field, helps determine the robustness of the reprogramming process. GFP-positive colonies that cannot be maintained or validated through in vitro assays represent partially reprogrammed iPSCs, which exhibit deficiencies in developmental potential and, thus, are not considered bona fide iPSCs. To address this, we employ a pluripotency validation strategy customary in the stem cell community. Initially, we establish a clonal iPSC line through three rounds of colony picking, ensuring that the iPSCs being tested originated from a single successfully reprogrammed cell. If these clonal iPSCs can be stably maintained in vitro, we proceed to evaluate the expression of core pluripotency markers using both ICC staining and RT-qPCR analysis. To assess the developmental potential of the clonal iPSCs, we perform tri-lineage differentiation in vitro by culturing them in differentiation media and examining the expression of markers representing three lineages. Additionally, we conduct in vivo assays by microinjecting the clonal iPSCs into blastocysts or morulae to generate chimeric mice. This rigorous in vivo experiment demonstrates the ability of these clonal iPSCs to contribute to the development of viable and healthy mice, confirming their status as bona fide pluripotent stem cells. We hope that these clarifications provide a clearer understanding of our experimental procedures. By no means we only look at the activation of reporter genes which could indeed be noisy. We use a panel of laborious and time-consuming assays to show that the iPSCs generated with unicellular Sox are indeed self-renewing and pluripotent stem cells. We have amended the text and legends to make this clearer in the revised manuscript.



Reviewer Figure 2. Schematic illustration of iPSC reprogramming and pluripotency validation.

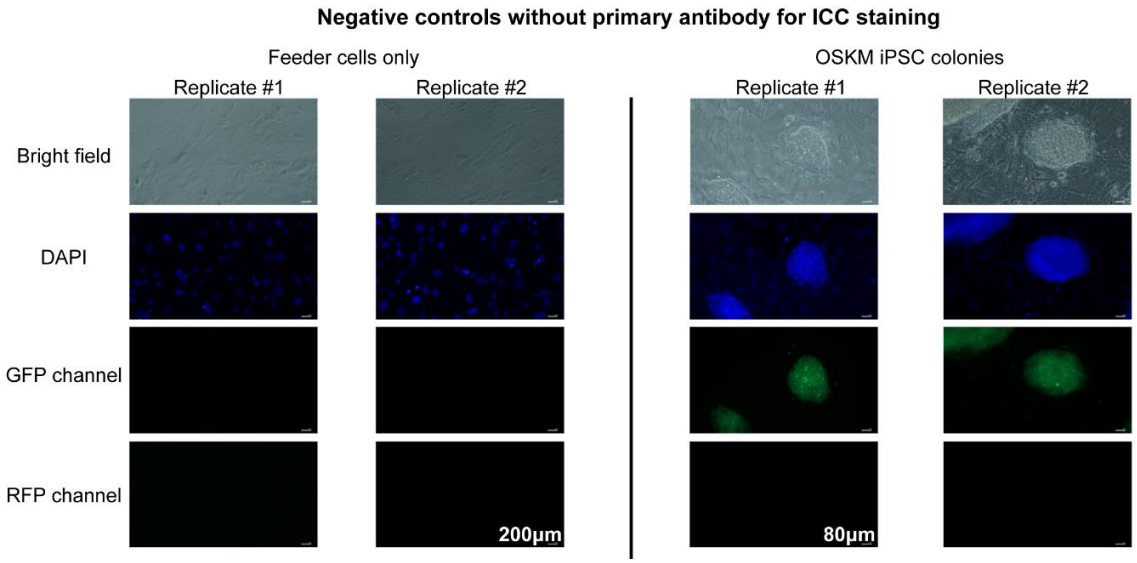
Coincident with this, the authors imply that pluripotent stem cells had a single origin in the stem animal (Fig 1A) and that these cells are homologous across metazoans. Even if there were support for the homology of pluripotency across animals (which there is not, see: Srivastava et al., 2021 Ann Rev Cell Dev Biol) the OSKM genes that drive embryonic pluripotency in mammals are certainly not conserved across animals. The authors point to cnidarians and planarians as examples of other animal taxa that have clear pluripotent stem cells but both of these taxa lack orthologs of Oct4 and Nanog and the unusual pluripotent stem cells of Hydra seem to be specific to that lineage of cnidarians, rather than an ancestral trait of the phylum (see: Chapman et al., 2010 supplementary information). The authors erroneously point to several citations to suggest an implicit relationship between sox/pou genes in cnidarians and placozoans but none of the papers they cite actually demonstrate evidence of sox/pou co-expression (rather, they just demonstrate both gene families are present in these taxa). Furthermore, recent evidence from another cnidarian (Nematostella) suggests SoxC is the multipotent stem cell regulator upstream of the more restricted SoxB genes and there's no evidence pou genes are so expressed with SoxC or soxB genes in this system. Suggesting there is an (undemonstrated) ancient interaction between sox and pou factors is misleading to readers that may be unfamiliar with the biology of these early-diverging metazoan lineages.

Response: The reviewer is right pointing out that the homology of adult pluripotent stem cells cannot be taken for granted across animal evolution, yet we do not propose such a scenario in the manuscript. However, pluripotency is inherent to development, any animal that undergoes embryonic development will have a stage where cells are pluripotent, therefore, it is certain that the last animal common ancestor possessed a pluripotent stem cell, albeit as a transient state. If that was regulated by SoxB or any POU member, remains a hypothesis. Still, our findings that unicellular Sox genes can replace Sox2 in mouse stem cell generation, and that they do that by interacting with Oct4, imply that this is a conserved biochemical feature, and that these partnerships between Sox and POU might be ancestral. What we propose is that Sox and POU might be important to establish stem cells across animals, not that they are the ancestral regulators of pluripotency, a concept we restrict to our iPSC mouse experiments. We have made changes throughout the text to make this distinction clearer. Not all stem cells are pluripotent, and as the reviewer mentions, neural progenitor stem cells are frequently marked by SoxB members, including cnidarians such as *Nematotsella vectensis* and *Hydractinia symbiolongicarpus*. Noteworthy, in the recent study taking advantage of single cell data on *Nematostella* development, it is mentioned that Sox3 is upstream of SoxC in the cascade of progenitors of NPCs, and Sox3 is a member of SoxB, quoting the study:
"*Regulatory genes upstream of NPCs include the HMG family member Sox3, zinc-fingerproteinOZF-like, E3SUMO-protein ligase NSE2-like, and 3 paralogs of the bHLH family of transcriptional inhibitors of the HES family (''putative stem cells'"*
If *Nematostella* Sox3 is a marker of stem cells in this species is something we now mention in discussion, yet it will need to be validated experimentally on that system. Given that Nematostella encodes many paralogues of SoxB, it is only expected that not all its members are regulating stem cells. In the hydrozoan *Hydractinia*, i-cells (the adult pluripotent stem cells) also express a SoxB member, yet this might be a convergent evolution of an adult pluripotent stem cell in this lineage, as argued by the reviewer. Interestingly, also in *Hydractinia* there are reports of a POU factor being important for stem cell maintenance (PMID: 21610024), which would lend support for our hypothesis. In placozoans, SoxB (and SoxC) members are expressed in the peptidergic progenitor stem cells (Najle et al 2023 Cell). In planarians, SoxB members are expressed in neoblasts (Wolfswinkel et al Cell Stem Cell 2014). In the demosponge *Amphimedon queenslandica*, SoxB and POU6 are co-expressed in archaeocytes (the pluripotent stem cells in sponges) and pinacocytes (Sogabe et al 2019 Nature, Sebé-Pedrós et al 2018 Nat Ecol Evol). Perhaps these are convergent events of SoxB deployment to stem cells. We agree that the role of POU factors in these stem cells and a possible Sox/POU partnership needs to be studied in further detail. We now mention these potential alternative scenarios, and highlight more clearly what are the differences between what

we know of pluripotency networks in vertebrates and what is known for other invertebrates.

In addition to the conceptual concerns outlined above, there is generally insufficient detail to evaluate much of the data presented in this manuscript. The authors provide the number of colonies that express GFP in their in vitro experiments rather than a percentage of total colonies, which makes these results hard to discriminate from noise. They also provide some measure of “efficiency” in a heat map but do not explain how this metric was calculated. The only evidence of actual pluripotency is demonstrated by immunofluorescence (Fig 2D,E) but these figures include neither positive controls (with mouse Sox2) nor negative controls (without primary antibody). For cooperativity assays, the authors fail to explain the axes in their figures (Fig 1F,G) and while they say binding was “highly specific” it’s not clear how this was determined. Again, it seems a different control would have been more informative, for example what would Cic look like in these same assays? The authors clearly have a lot of experience with cell culture but if this manuscript is marketed for a broad readership, the figures and methods will need be presented with much greater transparency.

Response: Thank you for your comments. For the quantifications, we changed to show the iPSC reprogramming efficiency by normalizing the colony numbers to the positive control, mSox2. As for the heat map, we initially explained the method for calculation in our supplementary tables. But to avoid confusion, we changed the heat map from showing the reprogramming efficiency to showing the success rate for generating GFP+ colonies and establishing stable iPSC lines in different biological replicates. For the immunofluorescence staining assays, we added the positive controls (iPSCs generated by mSox2 plus OKM, OSKM iPSCs) (Figure 2D), negative control #1 (staining for the feeder cells without any iPSCs seeded), and negative control #2 (staining for OSKM iPSCs without primary antibodies) (Reviewer Figure 3).



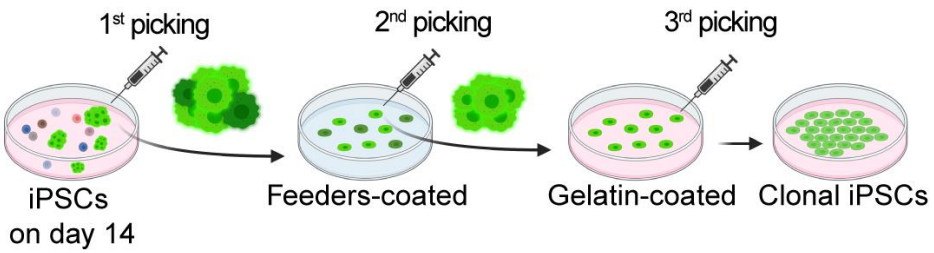
Reviewer Figure 3. Tests of negative controls for ICC staining by excluding primary antibody.

For the cooperativity assays, we have added additional figures on how the cooperativity factor (ω) is quantified and calculated (Figure S6A,B). The Specificity was evaluated in a separate assay called Specificity by Sequencing which evaluates specificity by evaluating binding with a library of sequences and calculating binding energy. Thanks for suggesting having the Cic as a control, we have added this as a control in iPSC reprogramming (new Figure S3A, B). We also purified the HMG DBD of Cic and evaluated its binding affinity (new Figure S2G,H) and cooperativity which we show in Rev Fig 4 below (as expected, it cannot substitute for Sox2 to make iPSCs or cooperate with Oct4)

Specific comments:
No citations are provided to support how the authors identified “key” residues in Sox and POU orthologs (Fig 1C and 5B)

Response: We have added references to reviews describing the structural basis for DNA binding and dimerization by Sox and POU (PMIDs: 29335749, 27521520) to the relevant figure legends.

Many of the methods are vague - What is a “round” of picking for data in Fig2 and S3? What constitutes an “experiment” when referring to replicates? What, specifically, do the authors consider to be “biological” vs “technical” replicates?



Reviewer Figure 4. Schematic illustration of three rounds of colony picking for establishing clonal iPSC lines.

Response: To establish clonal stem cell lines from iPSC reprogramming experiments, cells are sequentially isolated by so-called colony 'picking'. Proper pluripotent stem cells can make identical copies for indefinite periods even starting from single cells (self-renewal). The colony picking is done to ensure that the resulting iPSC lines are from a single independent reprogramming event derived from a genetically identical 'mother cell' or colony. Three "rounds" of picking means that we performed the colony-picking procedure three times. To provide further clarity, we have included a schematic illustration of the entire colony-picking procedure. Each round of colony picking involved selecting a colony, dissociating it into single cells, and replating the single-cell suspension to allow the cells to grow into new colonies. We then repeated the colony-picking process twice. By performing this procedure three times, we ensure clonality and the purity of the final iPSC cline.
Regarding the distinction between "biological" and "technical" replicates, we consider the use of the same batch of reagents, cells, and virus as technical replicates within an experiment. On the other hand, using different batches of MEF cells (taken from different mouse embryos), and virus represents biological replicates. This distinction allows us to assess the reproducibility and reliability of our experimental results from both technical and biological perspectives. We revised text and figure to clarify.

HMG and POU domains were validated with PFAM database but no indication of the significance cutoff (E-value) is provided.

Response: We have included these values in the methods section (e-value < 0.0001).

Fig 1C – do the logos indicated "unicellular holozoans" include M. brevicolis and S. rosetta or just the new taxa presented in this manuscript? It would be helpful if the authors could also include a logo for non-Sox HMG boxes for comparison.

Response: Yes, they do and we have updated Fig 1C with Sox-like sequences from other species (shown in Fig S2A). As suggested, we have also added the logo for TCF/LEF, a non-Sox HMG.

Fig 3C,D - The binding assay figures look identical for both taxa, how does this indicate different cooperativity? How does this translate into different abilities to induce stemness? Also the figures need to be explained more fully – what are the numbers and +/- signs at top indicating?

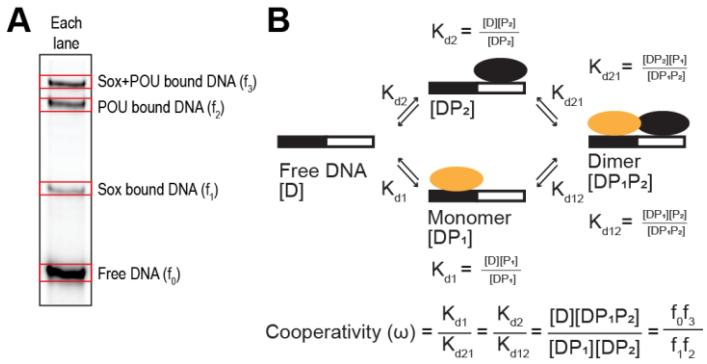
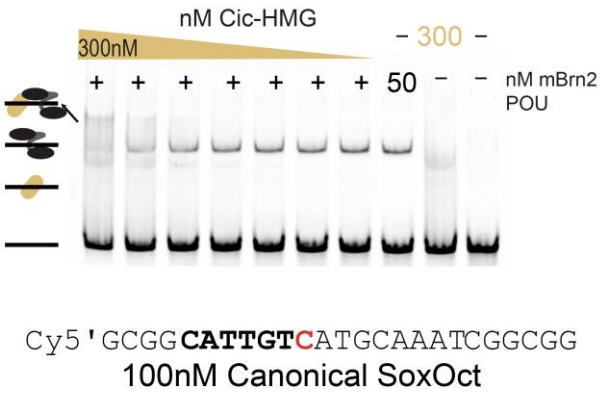


Figure S6A-B. (A) An illustration of DNA complex microstates present at each lane that were quantified for calculation of cooperativity. The cy5-labelled DNA migrated at different lengths on the native gel depending on how the proteins and DNA associate. Thus, the fractional contribution of the microstates of the free DNA (f0), Sox-DNA (f1), POU-DNA (f2) and ternary complex (f3) can be quantified. (B) Schematic diagram highlighting the approach to calculating the cooperativity of heterodimer formation on DNA as originally detailed in [cite Ng 2012]. Monomeric binding of either Sox or POU on each of their respective half sites followed by heterodimer complex formation by binding to the remaining site. Kd1 and Kd12 are the equilibrium dissociation constants for monomer or dimer formation. [D], [P], [DP1],[DP2] and [DP1P2] are the fractional concentrations of the free DNA, protein, monomer–DNA complexes and dimer-DNA complex, respectively. These concentrations can be estimated through the fractions (f0-f3 detailed in A). The cooperativity factor (ω) can then be calculated directly from these factional contributions.

Response: We have added additional explanatory figures in Fig S6A-B on how the cooperativity factor (ω) is calculated. The intensity of each band has to be accounted for and not just a direct comparison of the dimeric band. We have also changed the labels on the gels in Fig 3 for easier understanding and added additional information in the legends to describe the + and - signs on top that indicate the presence or absence of the protein. The SoxOct motif is present in pluripotency enhancers and higher cooperativity correlated with the capacity to induce pluripotency.

Fig 3E – what would the cooperativity of mouse/human Cic look like? What about sox from a sponge or cnidarian or planarian?



Reviewer Figure 5. Heterodimer EMSAs of Cic-HMG with mBrn2.

Response: Thank you for this interesting question. We've purified and tested Cic-HMG. As shown in FigS2G-H, Cic-HMG has a poor affinity with the Sox motif. Since heterodimer formation on the SoxOct element is dependent on binding on the Sox motif, we are not able to evaluate cooperativity on this motif as evidenced by the smeared bands in Rev Fig. 5. While what we found testing this is interesting, we have chosen not to include these in the manuscript. Regarding testing SoxB from other animals, these experiments are substantial as there are many species and paralogues that could be tested (e.g. just *Nematostella* has 5 SoxB paralogues). Since we do not think this would be highly informative (finding that a conserved family orthologue is capable of acting like Sox2) we have considered outside of the scope of this manuscript (testing the capacity of the unicellular Sox co-orthologues to perform the functions of one of its descendant paralogues).

Fig S1 – the previously identified sox genes from Mbrevicolis and Srosetta should be clearly labeled to match those sequences in the alignment in Fig S2. There are only two sox-like genes for each species in the alignment but there are three for Srosetta in the tree. (Same for Pigoraptor – sequences in tree should also be represented in the alignment.) Taxon abbreviations should be indicated. Some of the sequences are just indicated as "contig_XXX".
Table S1 is not provided.

Response: Thank you for spotting this, we have updated the figures to match the tree in Fig S1.

Fig S7A – ancestral state posterior probability plots – these are not explained at all. What do the probabilities correspond to - specific residues? Whole HMG binding domains? The authors need to provide the actual sequences that were cloned and expressed in cell culture and explain how fluorescence was quantified (Fig 4C SoxCEF looks fluorescent to me).

Response: We have added a more detailed description of what the ancestral state posterior probability plots show to the figure legend of Figure S9. Additionally, we have included the full sequences of the inferred ancestral HMG domains in the supplementary files. Please see our detailed explanation for the iPSC assays and a rigorous assessment of pluripotency provide above and in the revised text. Clonal SoxCEF lines could not be established for in these experiments.

REVIEWER COMMENTS

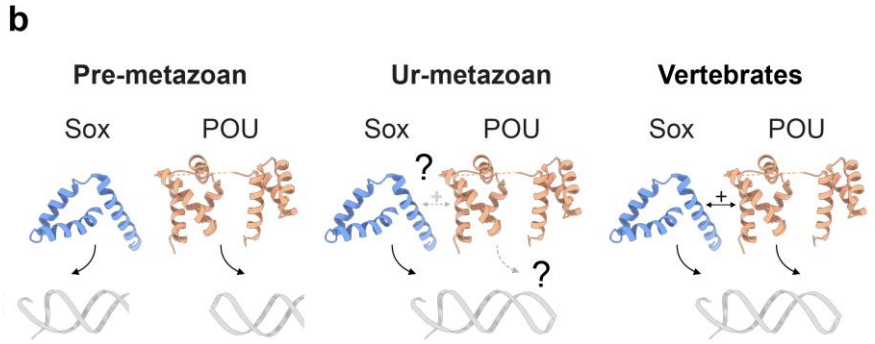
Reviewer #1 (Remarks to the Author):

The authors have extensively revised the manuscript, adding additional experiments and analyses, and removing many of the most (evolutionarily) speculative statements. This has resulted in a substantially improved manuscript.

I have no remaining major concerns, although I feel that the overall interpretation of the POU results, while improved and substantiated with the new POU3 experiment, is still biased towards trying to fit the vertebrate Sox-POU model into a distant ancestor. To do this thoroughly, given the lack of compelling resolution at critical nodes of the POU phylogenetic tree, the authors would need to do array of experiments with diverse ancestral POU family members. This is perhaps beyond the scope of this study. Instead, the authors should tone-down their extrapolations.

Response:

We have amended our summary figure 6b to clarify that it remains to be further examined if Sox and POU heterodimers outside vertebrates in particular in non-bilaterian animals.



Supplementary Figure 6b. The proposed model for the evolutionary innovation from pre-metazoans to ur-metazoan and vertebrates.

Reviewer #2 (Remarks to the Author):

The authors have fully addressed my questions and comments, and I am happy with the improved version of the manuscript.

Response: We are glad the reviewer is happy with our revision.

Reviewer #3 (Remarks to the Author):

The authors refer to the biochemical properties of Sox genes to regulate iPSCs as "already present" in choanoflagellates but this is a misrepresentation of the data. They have not shown this property to be shared by SoxB genes from any of non-bilaterian taxa (which would support their implication of continuity) and they imply that the properties exhibited by extant choanoflagellates necessarily represent the properties of these proteins in the choano/animal ancestor. This language needs to be modified.

Response: Whilst our data with ancestral Sox factors provide strong support for the argument that there is continuity along a trajectory from Ur-Sox towards animal SoxB, we agree that future work should systematically examine extant SoxB from a diverse set of taxa. We have adjusted the text and modified language accordingly stressing that we based our conclusions on extant unicellular Sox and ancestrally reconstructed versions.

Fig 1A still implies that stem cells are homologous and have a single origin in the common ancestor of animals, despite the claim that this is not implied in the text. The authors still use the generalized

language of "stem cells" when describing non-bilaterians but do only a cursory job of explaining that these cells do not fit the pattern of an ancestral SoxB/POU relationship.

Response: We have removed the stem cell symbol from Figure 1a to acknowledge the possibility that across animals might have emerged convergently. We also adjust the language regarding the description of stem cells in non-bilaterians.

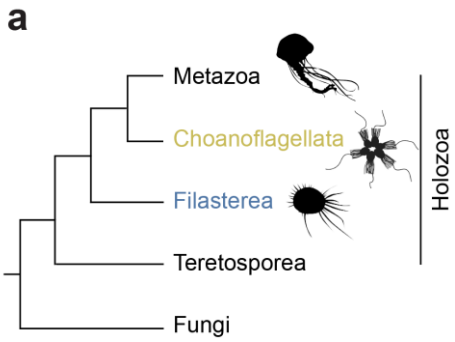


Figure 1a. Phylogeny of holozoans.