<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Lynch, Vincent J.; Bedoya-Reina, Oscar C.; Ratan, Aakrosh; Sulak, Michael; Drautz-Moses, Daniela I.; Perry, George H.; Miller, Webb; Schuster, Stephan C.</td>
</tr>
<tr>
<td><strong>Date</strong></td>
<td>2015</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10220/38768">http://hdl.handle.net/10220/38768</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>).</td>
</tr>
</tbody>
</table>
Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic

Graphical Abstract

Highlights
- Complete genomes of three Asian elephants and two woolly mammoths were sequenced
- Mammoth-specific amino acid changes were found in 1,642 protein-coding genes
- Genes with mammoth-specific changes are associated with adaptation to extreme cold
- An amino acid change in TRPV3 may have altered temperature sensation in mammoths

Authors
Vincent J. Lynch, Oscar C. Bedoya-Reina, Aakrosh Ratan, ..., George H. Perry, Webb Miller, Stephan C. Schuster

Correspondence
vjlynch@uchicago.edu (V.J.L.), webb@bx.psu.edu (W.M.)

In Brief
Lynch et al. sequence complete genomes from three Asian elephants and two woolly mammoths and identify amino acid changes unique to woolly mammoths. Woolly-mammoth-specific amino acid changes underlie cold-adapted traits in mammoths, including small ears, thick fur, and altered temperature sensation.
Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic

Vincent J. Lynch,†,* Oscar C. Bedoya-Reina,‡,∥ Aakrosh Ratan,‡,∥ Michael Sulak,† Daniela I. Drautz-Moses,†,∥ George H. Perry,∥ Webb Miller,* and Stephan C. Schuster‡,∥

†Department of Human Genetics, The University of Chicago, 920 East 58th Street, CLSC 319C, Chicago, IL 60637, USA
‡Center for Comparative Genomics and Bioinformatics, Pennsylvania State University, 506B Wartik Lab, University Park, PA 16802, USA
∥Departments of Anthropology and Biology, Pennsylvania State University, 513 Carpenter Building, University Park, PA 16802, USA
*Correspondence: vlynch@uchicago.edu (V.J.L.), webb@bx.psu.edu (W.M.)
http://dx.doi.org/10.1016/j.celrep.2015.06.027
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

SUMMARY

Woolly mammoths and living elephants are characterized by major phenotypic differences that have allowed them to live in very different environments. To identify the genetic changes that underlie the suite of woolly mammoth adaptations to extreme cold, we sequenced the nuclear genome from three Asian elephants and two woolly mammoths, and we identified and functionally annotated genetic changes unique to woolly mammoths. We found that genes with mammoth-specific amino acid changes are enriched in functions related to circadian biology, skin and hair development and physiology, lipid metabolism, adipose development and physiology, and temperature sensation. Finally, we resurrected and functionally tested the mammoth and ancestral elephant TRPV3 gene, which encodes a temperature-sensitive transient receptor potential (thermoTRP) channel involved in thermal sensation and hair growth, and we show that a single mammoth-specific amino acid substitution in an otherwise highly conserved region of the TRPV3 channel strongly affects its temperature sensitivity.

INTRODUCTION

Woolly mammoths (Mammuthus primigenius), perhaps the most charismatic of the extinct Pleistocene megafauna, have long fascinated humans and have become emblems of the last ice age. Unlike the extant elephantids, which live in warm tropical and subtropical habitats, woolly mammoths lived in the extreme cold of the dry steppe-tundra where average winter temperatures ranged from −30°C to −50°C (MacDonald et al., 2012). Woolly mammoths evolved a suite of adaptations for arctic life, including morphological traits such as small ears and tails to minimize heat loss, a thick layer of subcutaneous fat, long thick fur, and numerous sebaceous glands for insulation (Repin et al., 2004), as well as a large brown-fat deposit behind the neck that may have functioned as a heat source and fat reservoir during winter (Boeskorov et al., 2007; Fisher et al., 2012). They also likely possessed molecular and physiological adaptations in circadian systems (Bloch et al., 2013; Lu et al., 2010) and adipose biology (Ji et al., 2014; Nelson et al., 2014), similar to other arctic-adapted species. Mammoths diverged from Asian elephants (Elephas sp.) ~5 Ma (Rohland et al., 2007) and likely colonized the steppe-tundra 1~2 Ma (Debruyne et al., 2009), suggesting that their suite of cold-adapted traits evolved relatively recently (Figure 1).

Identifying the genetic changes that underlie morphological differences between species is daunting, particularly when reconstructing how the genotype-phenotype map diverged in non-model or, especially, extinct organisms. Thus, while the molecular bases of some phenotypic traits have been identified, these studies generally are limited to a few well-characterized genes and pathways with relatively simple and direct genotype-phenotype relationships (Chan et al., 2010; Hoekstra et al., 2006; Lang et al., 2012; Smith et al., 2013; Storz et al., 2009). Previous structural and functional studies, for example, have shown that amino acid polymorphisms in the woolly mammoth hemoglobin βδ fusion gene (HBB/HBD) reduce oxygen affinity (Campbell et al., 2010; Yuan et al., 2011, 2013), whereas amino acid polymorphisms in both the woolly mammoth and Neandertal melanocortin 1 receptor (MC1R) genes were hypomorphic compared to the ancestral alleles (Lalueza-Fox et al., 2007; Römler et al., 2006). Most traits, however, have complex genotype-phenotype relationships with phenotypic divergence arising through the accumulation of numerous variants of small individual effects rather than one or a few mutations of large effect. Thus, candidate gene studies are poorly suited for forward genetic-based approaches to trait mapping, and the genetic changes that
underlie woolly mammoth adaptations to the arctic are almost entirely unknown.

Whole-genome sequencing (WGS) is an invaluable tool for exploring the genetic origins of phenotypic differences between species, because one can identify fixed and polymorphic variants across the genome without respect to a-priori-defined genes and pathways. However, distinguishing functional from nonfunctional variants in WGS data can be difficult (Cooper and Shendure, 2011). To determine genetic changes that underlie cold-adapted traits in woolly mammoths, we sequenced the genomes of three Asian elephants and two woolly mammoths to high coverage, and we functionally annotated fixed, derived amino acid and loss-of-function (LOF) substitutions in woolly mammoths. We found that genes with woolly mammoth-specific substitutions were enriched in functions related to circadian biology, skin, hair, and sebaceous gland development and physiology, lipid metabolism, adipose development and physiology, and temperature sensation. These data provide mechanistic insights into the causes of morphological evolution, and define a set of likely causal variants for future study of woolly mammoth-specific traits.

RESULTS AND DISCUSSION

Genome Sequencing, Assembly, and Annotation

We generated Illumina sequence data for two woolly mammoths that died ~20,000 and ~60,000 years ago (Gilbert et al., 2007, 2008; Miller et al., 2008), including individuals from the two major lineages of woolly mammoths, clade I (individual M4) and clade II (M25), which are estimated to have diverged ~1.5 Ma (Miller et al., 2008), and three extant Asian elephants (Elephas maximus). We aligned sequencing reads to the genome assembly for the African Savannah elephant (Loxodonta africana), resulting in non-redundant average sequence coverage of ~20-fold for each mammoth and ~30-fold for each Asian elephant (Figure S1). We identified ~33 million putative single-nucleotide variants (SNVs) among the three elephantid species (see Experimental Procedures for details), including ~1.4 million nucleotide variants fixed for the derived allele in the two mammoths, but for the ancestral allele in the African and Asian elephants. Among the variants were 2,020 fixed, mammoth-derived amino acid substitutions in 1,642 protein-coding genes and 26 protein-coding genes with premature stop codons (putative LOF substitutions).

Functional Consequences of Woolly-Mammoth-Specific Amino Acid Substitutions

We used several complementary approaches to infer the putative functional consequences of mammoth-specific amino acid substitutions, including classifying substitutions based on their BLOSUM80 exchangeabilities (Henikoff and Henikoff, 1992), predicted functional consequences based on PolyPhen-2 (Adzhubei et al., 2010, 2013), and the inter-species conservation of sites at which substitutions occurred, as well as identifying Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Kanehisa and Goto, 2000) and mouse knockout (KO) phenotypes (Blake et al., 2014) enriched among protein-coding genes with fixed, derived amino acid substitutions in the wooly mammoth. Finally, we manually selected gene-pathway and gene-phenotype associations for further study according to the following two criteria: (1) the richness of literature supporting the role of each gene in specific pathways and phenotypes; and (2) the exchangeability, PolyPhen-2 score, and strength of sequence conservation at sites with mammoth-specific substitutions.

We found that genes with fixed, derived woolly mammoth substitutions were enriched for 40 KEGG pathways and 859 mouse KO phenotypes, at a false discovery rate (FDR) ≤ 0.10 (Figure 2A). Significantly enriched KEGG pathways included circadian rhythm - mammal (enrichment [E] = 6.71, hypergeometric p = 2.7 × 10⁻³, FDR q = 0.02), fat digestion and absorption (E = 4.01, hypergeometric p = 7.9 × 10⁻³, FDR q = 0.05),
complement and coagulation cascades ($E = 4.28$, hypergeometric $p = 5.0 \times 10^{-3}$, $FDR q = 6.7 \times 10^{-5}$), and metabolic pathways ($E = 8.39$, hypergeometric $p = 2.2 \times 10^{-3}$, $FDR q = 1.6 \times 10^{-5}$) (Table S1). Enriched KO phenotypes included decreased core body temperature ($E = 4.15$, hypergeometric $p = 8.0 \times 10^{-3}$, $FDR q = 7.2 \times 10^{-5}$), abnormal brown adipose tissue morphology ($E = 2.99$, hypergeometric $p = 1.4 \times 10^{-3}$, $FDR q = 4.0 \times 10^{-3}$), abnormal thermal nociception ($E = 2.25$, hypergeometric $p = 5.4 \times 10^{-3}$, $FDR q = 0.05$), abnormal glucose homeostasis ($E = 1.46$, hypergeometric $p = 2.6 \times 10^{-3}$, $FDR q = 3.2 \times 10^{-3}$), and many body mass-/weight-related phenotypes (Figure 2B).

We also inferred the functional significance of fixed, derived LOF substitutions in woolly mammoth genes. We identified a single KEGG term enriched among the genes with LOF substitutions, fat digestion and absorption ($E = 127.64$, hypergeometric $p = 1.0 \times 10^{-3}$, $FDR q = 1.0 \times 10^{-3}$), and 48 KO terms enriched among these genes at an $FDR \leq 0.10$ (Figure 3A). Enriched KO terms were almost exclusively related to cholesterol, sterol, triglyceride, and lipid homeostasis and metabolism (Figure 3B), such as decreased circulating cholesterol level ($E = 33.15$, hypergeometric $p = 5.7 \times 10^{-5}$, $FDR q = 4.3 \times 10^{-3}$), decreased sterol level ($E = 30.15$, hypergeometric $p = 7.6 \times 10^{-5}$, $FDR q = 4.3 \times 10^{-3}$), and abnormal circulating lipid level ($E = 30.15$, hypergeometric $p = 7.6 \times 10^{-5}$, $FDR q = 4.3 \times 10^{-3}$).

**Substitutions in Genes Associated with the Mammoth Body Plan**

Woolly mammoths evolved a suite of morphological adaptations to life in the extreme cold, including small ears and tails, a long thick coat, and, unlike other elephants, numerous large sebaceous glands, which are thought to have helped repel water and improve insulation (Repin et al., 2004). Woolly mammoths also evolved a characteristic set of skeletal traits, including a high, domed skull with dorsally expanded parietals; an anteroposteriorly compressed skull; and a sloping back. Consistent with mammoth-specific amino acid changes contributing to these traits, we found that genes with mammoth-specific substitutions were enriched in KO phenotypes such as abnormal tail morphology ($E = 1.71$, hypergeometric $p = 2.0 \times 10^{-3}$, $FDR q = 2.2 \times 10^{-3}$), abnormal tail bud morphology ($E = 5.06$, hypergeometric $p = 3.0 \times 10^{-3}$, $FDR q = 3.2 \times 10^{-2}$), small tail bud ($E = 18.00$, hypergeometric $p = 4.1 \times 10^{-2}$, $FDR q = 1.6 \times 10^{-2}$),
abnormal ear morphology (E = 1.60, hypergeometric p = 9.0 \times 10^{-3}, FDR q = 6.4 \times 10^{-2}), cup-shaped ears (E = 5.06, hypergeometric p = 3.0 \times 10^{-3}, FDR q = 1.4 \times 10^{-2}), and abnormal sebaceous gland morphology (E = 2.33, hypergeometric p = 8.0 \times 10^{-2}, FDR q = 6.3 \times 10^{-2}), including substitutions in three genes leading to enlarged sebaceous glands. We also found numerous enriched KO phenotypes that affected the skeleton, including domed cranium (E = 2.26, hypergeometric p = 1.3 \times 10^{-2}, FDR q = 8.3 \times 10^{-2}), abnormal parietal bone morphology (E = 2.66, hypergeometric p = 2.6 \times 10^{-2}, FDR q = 1.9 \times 10^{-2}), and short snout (E = 2.39, hypergeometric p = 2.0 \times 10^{-2}, FDR q = 2.3 \times 10^{-2}).

Previous studies have found hypomorph polymorphisms in the woolly mammoth MC1R associated with reddish fur color (Römpfler et al., 2006), but these variants may have been relatively rare in mammoth populations (Workman et al., 2011). Thus, variants in other genes also may have contributed to coat color variability in mammoths, which varied from blonde to orange to nearly black (Valente, 1983). We identified 38 genes with mammoth-specific amino acid changes associated with abnormal coat/hair morphology in KO mice, including derived substitutions in eight genes specifically associated with diluted coat color. We also found that the expression of genes with fixed, derived woolly mammoth substitutions were enriched in hair root sheath (hypergeometric p = 0.006), coat hair follicle (hypergeometric p = 0.013), hair follicle (hypergeometric p = 0.016), skin (hypergeometric p = 0.018), and hair outer root sheath (hypergeometric p = 0.018).

**Substitutions in Genes Associated with Circadian Biology**

Organisms living at high latitudes in the arctic experience long periods of darkness during the winter and near constant light in the summer, which prevents polar-adapted species from utilizing daily light-dark cycles to entrain their circadian clocks. Svalbard reindeer (*Rangifer tarandus platyrhynchus*), for example, have lost functioning circadian clocks and circadian rhythmicity in *PER2* and *BMAL1* (*ARNTL*) expression (Lu et al., 2010). Moreover, several other arctic species also are known to have derived circadian systems (Bloch et al., 2013), and we observed that several enriched KO and KEGG terms were related to circadian biology, motivating us to explore circadian genes in greater detail.

Fixed, derived mammoth-specific amino acid substitutions occurred in eight genes associated with circadian biology, including those that play central roles in maintaining normal circadian rhythms and entraining the circadian clock to external stimuli such as temperature. *HRH3* and *PER2* KO mice, for example, have abnormal circadian temperature homeostasis (Shiromani et al., 2004; Toyota et al., 2002). *PER2* directly mediates the early adaptive response to shifted temperature cycles and coordinates adaptive thermogenesis by synchronizing *UCP1* expression and activation in brown adipose tissue (Chapuis et al., 2013; Saini et al., 2012). Similarly, neuronal histamine receptors regulate circadian energy homeostasis through *UCP1* expression in brown adipose tissue. *HRH1* KO mice, for example, have abnormal circadian rhythms and abnormal
Substitutions in Genes Associated with Insulin Signaling, Lipid Metabolism, and Adipose Biology

The enrichment of genes with derived amino acid (Figure 2) and LOF substitutions (Figure 3) in woolly mammoths that function in lipid metabolism, adipose development, and physiology suggests modifications of these processes may have played an important role in the evolution of woolly mammoths and adaptation to arctic life. We identified 54 genes with fixed, derived amino acid substitutions and KO phenotypes that affect adipose tissue, including phenotypes that alter both the location and abundance of white and brown fat deposits throughout the body. Among the genes with woolly mammoth-specific substitutions were the leptin receptor (LEPR); DLK1 (also known as preadipocyte factor 1), an epidermal growth factor repeat-containing transmembrane protein that regulates adipocyte differentiation; the growth hormone receptor (GHR); and corticotropin-releasing hormone (CRH). We also identified 39 genes with KO phenotypes that affect insulin signaling, and found that genes with mammoth-specific amino acid substitutions were enriched in several KO phenotypes related to insulin signaling, including abnormal circulating insulin level (E = 1.82, hypergeometric p = 1.0 × 10^{-3}, FDR q = 1.5 × 10^{-2}), insulin resistance (E = 2.23, hypergeometric p = 3.0 × 10^{-3}, FDR q = 3.5 × 10^{-2}), and impaired glucose tolerance (E = 1.91, hypergeometric p = 4.0 × 10^{-3}, FDR q = 3.7 × 10^{-2}).

Substitutions in Temperature-Sensitive Transient Receptor Potential Channels

The most intriguing mouse KO phenotype enriched among genes with woolly mammoth-specific amino acid changes was abnormal thermal nociception (13 genes). For example, we identified woolly mammoth-specific amino acid changes in five temperature-sensitive transient receptor potential (thermoTRP) channels (Figure 4A) that sense noxious cold (TRPM8) (Bautista et al., 2007; Knowlton et al., 2010; Vriens et al., 2014), innocuous warmth (TRPV3 and TRPV4) (Chung et al., 2004; Smith et al., 2002; Vriens et al., 2014; Xu et al., 2002), or noxious cold or heat depending on species (TRPA1) (Chen et al., 2013; Karashima et al., 2009) or that are heat sensitive but not known to be involved in temperature sensation (TRPM4) (Bautista et al., 2007; Knowlton et al., 2010; Vriens et al., 2014). We also identified a mammoth-specific amino acid change in PIRT, a small phosphoinositide-binding protein that functions as a regulatory subunit of TRPM8 and the noxious heat sensor TRPV1 (Kim et al., 2008; Tang et al., 2013).

To infer the putative consequences of woolly mammoth-specific amino acid substitutions in thermoTRPs, we generated structural models of the ancestral Asian elephant/mammoth (AncGajah; Figure 1A) and ancestral mammoth (AncMammoth; Figure 1A) TRPA1, which mediates nociceptive (Karashima et al., 2009; Kwan et al., 2006; Vizin et al., 2015) and vascular responses to noxious cold (Aubdool et al., 2014) as well as generally potentiating responses to noxious stimuli (del Camino et al., 2010), and TRPV4, which mediates autonomic and behavioral responses to cold (Vizin et al., 2015). We found that the elephantid TRPA1 and TRPV4 proteins were predicted to adopt the common TRP channel structure, which is composed of a
series of amino terminal ankyrin repeats (ARD), separated by a membrane proximal domain (MPD) from the six transmembrane helices (S1–S6) that form the ion-permeable pore in tetrameric channels (Figure 4B). We found that the TRPA1 R1031T substitution occurred in an unstructured loop between the TRP-like domain and the C-terminal coiled-coil domain (Figures 4C and 4D), which is predicted to alter the electrostatic surface by reducing the local positive charge (Figure 4E). The mammoth-specific TRPV4 V658I substitution occurred at the first site in the S6 helix (Figure 5A), which is part of the outer pore region important for activation of the channel in response to heat (Figure 5B). Indeed, we found that site 658 is located within a cluster of sites that mediate heat activation in the related TRPV3 channel (Grandl et al., 2008), and it is homologous to a site in TRPV3 that adopts temperature-dependent conformations (Figures 5C and 5D; Kim et al., 2013). Site 658 also mediates the interaction between TRPV channels and the agonist vanillotoxin DkTx (Figure 5E; Cao et al., 2013; Liao et al., 2013). These data suggest the mammoth-specific R1031T and V658I substitutions may have affected the gating dynamics in TRPA1 and TRPV4, respectively.

**Thermal Tuning of the Woolly Mammoth Temperature Sensor TRPV3**

To explore the consequences of mammoth-specific amino acid changes in thermoTRPs in greater detail, we focused on TRPV3, which functions in a variety of processes including warm temperature sensation (>33°C) through ATP-dependent signaling between epidermal keratinocytes, which express TRPV3, and local sensory neurons, which do not (Chung et al., 2004; Mandadi et al., 2009; Peier et al., 2002; Vandewauw et al., 2013); regulating hair growth through transforming growth factor a (TGF-α)/epidermal growth factor receptor (EGFR) signaling (Cheng et al., 2010; Imura et al., 2007; Smith et al., 2002; Xu et al., 2002); and differentiation of adipocytes (Cheung et al., 2015). We found that the mammoth-specific substitution in TRPV3 (N647D) occurred at a well-conserved site (Figure S2) in the outer pore loop (Figures 6A and 6B) and was fixed for the derived asparagic acid in seven woolly mammoths we tested by PCR amplification and Sanger sequencing (Figure S3). Remarkably, a previous high-throughput mutagenesis screen found that mutations at site 647 abolished heat sensitivity in mouse TRPV3 (Grandl et al., 2008), suggesting the N647D substitution affects thermosensation by mammoth TRPV3.

We inferred the structural consequences of the N647D substitution by generating homology models of the AncGajah and AncMammoth TRPV3 protein and tetrameric channel (as described above). We found that the TRPV3 models were structurally very similar to the TRPV1 reference structure on which they were based (root-mean-square deviation [RMSD] = 1.93–1.99 Å; Figure S4), particularly in the α helices that form the pore of the tetrameric channel, suggesting that these are realistic models. In the TRPV1 structure, hydrogen bonds between residues in the pore loop are thought to maintain the outer pore in a non-conductive conformation in the closed state;
conformational changes in the pore helix and pore loops disrupt these local hydrogen bonds to facilitate gating and widening of the selectivity filter upon channel activation (Cao et al., 2013; Liao et al., 2013). Our structural model suggests that the carbonyl oxygen of the ancestral N647 residue forms a hydrogen bond with the neighboring side chain of Q645, whereas in the AncMammoth structure these hydrogen bonds are replaced by a pair of hydrogen bonds between D647 and K610, potentially impeding full opening of the channel (Figure 6C). Superimposed AncGajah (red) and AncMammoth (blue) pore regions in the open conformation. Only diagonally opposed subunits are shown. Site 647 is shown as spheres and sites G637 and I673 are sticks. (Left) Predicted pores formed by the AncGajah (red) and AncMammoth (blue) TRPV3 channels are shown as space-filling spheres. (Right) The diameter of the AncMammoth pore relative to the diameter of the AncGajah pore at the narrowest point in the selectivity filter (site G637) and lower gate (site I673) is shown.

(F) Fluo-4 fluorescence intensity in response to increases in temperature in HEK293 cells transiently transfected with expression constructs for AncGajah (red) and AncMammoth (blue) TRPV3 relative to non-transfected cells (gray). Curves are shown as background-subtracted relative intensity, mean ± SEM (n = 6).

To functionally characterize the effects of the mammoth-specific N647D substitution, we resurrected the AncMammoth and AncGajah TRPV3 genes and measured their temperature-dependent gating in transiently transfected HEK293 cells using Fluo-4 calcium flux assays (Aneiros and Dabrowski, 2009; Reubish et al., 2009). We found that both the AncMammoth and AncGajah TRPV3 proteins were expressed at similar levels (Figure S5I) and that the overall gating dynamics of channels were very similar. Both channels, for example, were activated at ~29°C, had half-maximum activities (T 50) at 33°C, and had maximal activities (T max) at 43°C (Figure 6F). The AncMammoth TRPV3 channel, however, was ~20% less active than the AncGajah channel at T max (Figure 6F), consistent with the predictions from our structural models that the AncGajah channel does not fully open upon stimulation. Both channels, however, were robustly activated in response to camphor (Figure S5J). These data are consistent with previous studies in mice that found mutations at site 647 affect temperature-dependent gating, but not channel opening, by chemical agonists (Grandl et al., 2008).
To test whether this substitution may have been positively selected in the woolly mammoth lineage, we assembled a dataset of TRPV3 genes from 64 diverse annectodes and used maximum likelihood methods to identify lineages (aBSREL) and codons with evidence of episodic (MEME) and pervasive (FEL) diversifying selection. While aBSREL identified a class of sites in mammoth with \( d_{N}/d_{S} > 1 \), the results were not significant (mean \( d_{N}/d_{S} = 10, \ p = 0.226 \)). MEME, however, found significant evidence for episodic diversifying selection at site 647 \( (d_{N}/d_{S} = 444.47, \ p = 0.037) \); although inferences of positive selection at specific branch-site combinations are inherently imprecise, the MEME model suggested the N647D substitution was positively selected in mammoths \( (P > 0.97, \ EBF > 1,000) \). In contrast, FEL inferred site 647 to evolve under strong purifying selection \( (d_{N}/d_{S} = 0.274, \ p = 0.044) \), indicating this site does not experience pervasive diversifying selection. These data suggest that, while TRPV3 genes and site 647 generally evolve under purifying selection, there is strong evidence that the N647D substitution was positively selected in the stem lineage of woolly mammoths.

Our observation that the mammoth TRPV3 protein is less active (hypomorphic) across a range of temperatures is particularly intriguing given its pleiotropic roles in temperature sensation, hair growth, and adipogenesis. TRPV3 KO mice, for example, have deficits in responses to innocuous and noxious heat and prefer colder temperatures than wild-type mice (Marics et al., 2014; Miyamoto et al., 2011; Moqrich et al., 2005; cf. Huang et al., 2011). TRPV3 activation also inhibits hair shaft elongation and induces the premature regression of hair follicles (Borbíró et al., 2011; Cheng et al., 2010), whereas TRPV3 KO mice have curly whiskers and wavy hair (Cheng et al., 2010). These data suggest that the hypomorphic mammoth TRPV3 may have phenocopied TRPV3-null mice and contributed to evolution of cold tolerance, long hair, and large adipose stores in mammoths.

Conclusions
Identifying the genetic changes that underlie morphological evolution is challenging, particularly in non-model and extinct organisms. We have identified genetic changes unique to woolly mammoths, some of which likely contributed to woolly mammoth-specific traits. Our results suggest that changes in circadian systems, insulin signaling and adipose development, skin development, and temperature sensation may have played important roles in the adaptation of woolly mammoths to life in the high arctic. Our identification of a hypomorphic woolly mammoth amino acid substitution in TRPV3 is particularly noteworthy given its pleiotropic roles in temperature sensation, hair growth, and adipose biology, suggesting that this substitution may have contributed to cold tolerance. Finally, the genomic data we have generated will be a useful resource for future studies to explore the genetic changes that underlie woolly mammoth morphology, physiology, and demography.

EXPERIMENTAL PROCEDURES

Genome Sequencing, Assembly, and Annotation
Details of the sequencing protocol are given in the Supplemental Experimental Procedures. Sequences from the three Indian elephant samples were aligned to the reference genome from the African elephant (loxAfr3) using the Burrows Wheeler Aligner (BWA (Li and Durbin, 2010) with default parameters (BWA version 0.5.9-r16). The reads were subsequently realigned around putative indels using the Genome Analysis Toolkit (GATK (DePristo et al., 2011)) IndelRealigner (version 1.5-21-g979a84a), and putative PCR duplicates were flagged using the MarkDuplicates tool from the Picard suite (version 1.96).

For the two mammoth samples, we trimmed putative adaptor sequences and merged overlapping paired-end reads using available scripts (Kircher et al., 2012). We required an overlap of at least 11 nucleotides between the mates, and only pairs that could be merged were retained for subsequent analyses. The merged reads were aligned to the genome from the African elephant (loxAfr3) using BWA with default parameters, and only the mapped reads that were longer than 20 bp were retained for the subsequent SNP calls. The reads were realigned using the GATK IndelRealigner and putative PCR duplicates were flagged using MarkDuplicates, similar to the process described for the modern genomes. We also limited the incorporation of damaged sites into the variant-calling pipeline by hard-masking all sites that would be potentially affected by the characteristic ancient DNA patterns of cytosine deamination in single-stranded overhangs. This mask was applied to ten nucleotides on both ends of the merged reads from the ancient samples.

At about 33 million positions in the African elephant reference assembly, we detected a nucleotide difference from the reference in at least one of the five newly sequenced individuals. We call these positions SNVs; these were identified using SAMtools (Li et al., 2009) (version 0.1.19), which was applied with "-C50" to adjust the mapping quality of the reads with multiple mismatches. We did not call differences in regions where the reference base was unknown, and the calls were limited to regions that were covered at least four times and at most 250 times by the sequences in these samples.

We selected the SNVs where the two mammoths were identified as homologous for the variant nucleotide, whereas the three Asian elephants were homozygous for the Loxodontina africana reference nucleotide. Since the African elephant is thought to have diverged from the ancestor of Asian elephants and mammoths (Krause et al., 2006), we considered the fixed mammoth variant as derived (i.e., non-ancestral). We used the gene annotation for Loxodontina africana to identify putative variant amino acids. This information as well as gene ontology (GO) terms and gene models were obtained from the Ensembl database (Flicek et al., 2013).

We also wanted to provide each SNV with quality values that can help determine the robustness of an analysis to potential erroneous SNV calls. It was not clear to us how to define a single quality value that treats mammoths on an equal footing with Asian elephants, because of the lower coverage, shorter length, and decreased accuracy of the mammoth reads, so we annotated each SNV with a mammoth quality value and an Asian elephant quality value, which gave the Phred scaled probability of the alternate allele in the two mammoth samples and the three Asian elephants, respectively. Slightly under half of the ~33 million SNV calls have a mammoth quality value of at least 100, but, of the 2,046 putative fixed mammoth-specific non-synonymous differences (2,020 amino acid variants and 26 premature stop codons), 1,975 have a mammoth quality value of at least 100, which suggests to us that our conclusions are reasonably robust. In any case, the user can filter the putative SNVs as desired. Among these variants were five previously characterized mammoth-specific amino acid changes (Miller et al., 2008); however, the remaining previously identified changes failed to pass our stringent quality control or were excluded because of different data analysis pipelines. The genes in which replicated variants were identified are TMEM48, NF5E, MARIS, LRRC49, and PRMT7. Promoted by our observation that numerous thermoTRPs had mammoth-specific amino acid changes, we also manually annotated the mammoth TRPM8 locus by lowering our thresholds for SNP calling and identified four derived mammoth amino acid changes.

A table of all 2,046 fixed, mammoth-specific protein differences is freely available on the Galaxy server (Goecks et al., 2010; Bedoya-Reina et al., 2013; https://usegalaxy.org). The table has the following columns: (1) gene name; (2) reference amino acid; (3) position in the peptide sequence (base 1); (4) variant amino acid; (5) name of Ensembl transcript; (6) name of scaffold in the Loxodonta genome assembly; (7) position in the scaffold (base 0); (8) name of orthologous human chromosome; (9) human position; (10) BLOSUM90 exchangeability score; (11) PolyPhen-2 category ("benign,"

Cell Reports 12, 217–228, July 14, 2015 ©2015 The Authors
“possibly damaging,” “probably damaging,” or “unknown”); (12) PolyPhen-2 score; and (13) mammoth SNV quality value. Tables of the 33 million SNVs, Loxodonta/Ensembl-annotated genes, 170,274 SNVs in those protein-coding regions, and a complete command history for constructing the table of 2,046 differences (see above) are available at https://usegalaxy.org/r/woolly-mammoth.

Functional Inference of Mammoth-Specific Amino Acid Substitutions
We used Vlad (http://proto.informatics.jax.org/prototypes/vlad/) to mine the mouse KO phenotype data at Mouse Genome Informatics (http://www.informatics.jax.org) for the genes with mammoth-specific substitutions. Enriched GOs and KEGG pathways were identified with WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/). The results can be found in Table S2.

TRPA1 and TRPV4 Structure Modeling
To reconstruct the AncMammoth and AncGajah TRPA1 and TRPV4 protein sequences, we did the following: (1) included TRPA1 or TRPV4 genes from the genomes of two woolly mammoths, three Asian elephants, African elephant (loxAf3), West Indian manatee, hyrax (proCap1), lesser hedgehog tenrec (TENREC), and nine-banded armadillo (dasNov2); (2) aligned the translated sequences with MUSCLE (Edgar, 2004); (3) inferred the best-fitting model of amino acid substitution using the model selection module implemented in Datamonkey (Delport et al., 2010); and (4) used joint (Pupko et al., 2000), marginal (Yang et al., 1995), and sampled (Nielsen, 2002) maximum likelihood methods implemented in the ancestral state reconstruction (ASR) module of Datamonkey, incorporating a general discrete model of site-to-site rate variation with three rate classes and the species phylogeny. We found that the AncGajah TRPA1 and TRPV4 protein sequences were inferred with support of 1.0 across all sites under the joint, marginal, and sampled likelihood methods.

The AncMammoth and AncGajah TRPV4 protein structures were modeled using the recently published high-resolution cryo-electron microscopy (EM) structure of human TRPV1 in the closed and open states (Cao et al., 2013; Liao et al., 2013; Paulsen et al., 2015). Initial structural models of the AncMammoth and AncGajah TRPV4 proteins in the open and closed states were generated using I-TASSER (Roy et al., 2010; Zhang, 2008) and the experimentally determined structure of the TRPV1 channel in the closed (PDB: 3J5P) and open (PDB: 3J5Q) states as templates. Initial AncMammoth and AncGajah structural models were refined with ModRefiner (Xu and Zhang, 2011), using the TRPV1 channel in the closed (PDB: 3J5P) and open (PDB: 3J5Q) states as the reference structure. Similarly, we modeled the AncMammoth and AncGajah TRPA1 protein structures using the high-resolution cryo-EM structure of human TRPA1 (Cao et al., 2013; Liao et al., 2013; Paulsen et al., 2015) using I-TASSER and ModRefiner.

TRPV3 Ancestral Sequence Reconstruction and Gene Synthesis
To reconstruct the mammoth ancestral TRPV3 protein sequences, we did the following: (1) included TRPV3 genes from the genomes of two woolly mammoths, three Asian elephants, African elephant (loxAf3), West Indian manatee, hyrax (proCap1), lesser hedgehog tenrec (TENREC), and nine-banded armadillo (dasNov2); (2) aligned the translated sequences with MUSCLE (Edgar, 2004); (3) inferred the JTT model as the best-fitting (AIC = 7,255.71, cAIC = 7,256.05, BIC = 7,380.52) model of amino acid substitution using the model selection module implemented in Datamonkey (Delport et al., 2010); and (4) used joint (Pupko et al., 2000), marginal (Yang et al., 1995), and sampled (Nielsen, 2002) maximum likelihood methods implemented in the ASR module of Datamonkey, incorporating a general discrete model of site-to-site rate variation with three rate classes and the species phylogeny. We found that support for the reconstructions at site 647 in both ancestral sequences was 1.0 under joint, marginal, and sampled likelihoods.

The Asian elephant/mammoth (AncGajah) ancestral TRPV3 gene, including an amino terminal flag tag (MDYKDDDDK) with a Kozak sequence incorporated into the FLAG tag (ggccattAGG) and a four-residue glycine spacer (GGGG) amino (N)-terminal to the TRPV3 open reading frame, was synthesized by GeneScript using human codon usage and cloned into the mammalian expression vector pcDNA3.1(+). The ancestral mammoth gene (AncMammoth) was generated by using site-directed mutagenesis to introduce the N847D mutation into the AncGajah TRPV3 pcDNA3.1(+) construct. The sequence of both ancestors was verified with Sanger sequencing.

TRPV3 Structure Modeling
The AncMammoth and AncGajah protein structures were modeled using the recently published high-resolution cryo-EM structure of TRPV1 in the closed and open states (Cao et al., 2013; Liao et al., 2013). Initial structural models of the AncMammoth and AncGajah proteins in the open and closed states were generated using I-TASSER (Roy et al., 2010; Zhang, 2008) and the experimentally determined structure of the TRPV1 channel in the closed (PDB: 3J5P) and open (PDB: 3J5Q) states as templates. Initial AncMammoth and AncGajah structural models were refined with ModRefiner (Xu and Zhang, 2011), using the TRPV1 channel in the closed (PDB: 3J5P) and open (PDB: 3J5Q) states as the reference structure. The backbone atoms of the refined AncMammoth and AncGajah structural models in their closed and open conformations then were aligned to the pore tetramer of the TRPV1 structure. Pore radius was measured with MOLE 2.0 (http://mole.upol.cz).

TRPV3 Function Assays
We used the Flu-4 NW Calcium Assay Kit (Life Technologies) to determine the temperature response of the AncMammoth and AncGajah TRPV3 proteins. HEK293 cells (ATCC CRL-1573), were cultured in MEM supplemented with 10% (v/v) fetal bovine serum (FBS) in a 37°C humidity-controlled incubator with 10% CO2. HEK293 cells growing in 10-cm plates were transiently transfected at 80% confluence with 24 μg expression vector for the AncMammoth or AncGajah TRPV3 genes or empty pcDNA3.1(+) using Lipofectamine LTX+ (Life Technologies), using the standard protocol.

Then, 48 hr after transfection, cells were harvested by trypsinization, centrifuged, resuspended in Hank’s balanced salt solution (HBSS) and HEPES assay buffer containing 2.5 mM probenecid, and transferred to a 96-well plate (at 150,000 cells/well in 50 μl assay buffer). Temperature-dependent calcium influx was assayed using the Flu-4 NW Calcium Assay Kit (Molecular Probes) and a high-throughput qPCR-based assay (Aneiros and Dobrowski, 2009; Reubish et al., 2009). After an initial 30-min loading at 25°C, the temperature was raised from 15°C to 57°C in 2°C steps, and fluorescence was measured after 2 min at each temperature using a Bio-Rad CFX-96 real-time PCR machine. Flu-4 fluorescence was measured using channel 1 (Sybr/FAM). Flu-4 fluorescence of cells transfected with the AncMammoth or AncGajah TRPV3 genes was normalized by the Flu-4 fluorescence of empty pcDNA3.1(+) transfected controls. All experiments included six biological replicates and were repeated in four independent experiments.

Data Availability
Tables of the nucleotide and amino acid differences that we identified and a table of putative gene gains and losses are available at the Galaxy website, and collected at https://usegalaxy.org/r/woolly-mammoth, along with the table of putative fixed woolly mammoth-specific amino acids and the set of Galaxy commands that created it. Those data can be further analyzed by a suite of Galaxy tools designed specifically for these data types (Bedoya-Reina et al., 2013).

ACCESSION NUMBERS
The accession number for the Asian elephant and woolly mammoth sequence data reported in this paper is SRA: PRJNA281811 (Short Read Archive).

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cellrep.2015.06.027.

AUTHOR CONTRIBUTIONS
V.J.L. designed and led the experimental analysis of TRPV3. V.J.L., W.M., and O.C.B.-R. analyzed the elephantid sequences. A.R. identified sequence
variants. M.S. performed experiments on TRPV3. D.I.D.-M. performed PCR validations. G.H.P. provided insights about sequence analysis and experimental methods. S.C.S. and W.M. led the woolly mammoth sequencing project. V.J.L. and W.M. wrote the paper with input from the coauthors.

ACKNOWLEDGMENTS

Sequencing was supported by National Science Foundation award EAR-0921958 to S.C.S. We thank O. Ryder and S. Santosh for providing Asian elephant DNA samples and M. DeGiorgio for insights about woolly mammoth SNVs. We also thank S.L. Kosakovsky Pond for assistance in interpreting the TRPV3 selection tests results.

REFERENCES


