PHASESPACE DESIGN LAB / CURRENT INVESTIGATIONS

Can We Turn a Shrub into a Perennial Protein Crop?

9 NOVEMBER 2026 • VERSION 1.0 • RESEARCH IN PROGRESS

IN BRIEF

There is a tough, long-lived legume that has spent much of its introduced life holding back the wind. We have been wondering whether it could do something else: produce a useful harvest of edible seed protein year after year. Our first image was a Siberian peashrub with pods eight to ten inches long. The biology quickly became more complicated. A longer pod need not contain more seeds; more seeds need not produce more recoverable protein; and changing a plant's defenses can change what survives until harvest. We have examined the published genomes and chemistry, developed preliminary computational models, and identified possible CRISPR research directions. **We have not engineered a Caragana plant or demonstrated a new food crop.** This letter describes what the investigation has made visible and what evidence we still need.

A windbreak that does not keep its harvest

Caragana arborescens, usually called Siberian peashrub, is familiar in colder parts of North America as a hedge or windbreak. Its yellow flowers develop into small pods. A US Forest Service seed manual describes pods about **2.5–5 centimetres long**, commonly containing around six seeds. The shrub may take three to five years to become commercially seed-bearing. Then, as the pods ripen, they split and scatter the seeds. The practical collection window can be less than two weeks. [1]

That last detail matters as much as the size. A plant can produce perfectly good seeds and still be a poor crop if it distributes them before anyone can gather them. Our imagined eight-to-ten-inch pods, roughly 20–25 centimetres, are deliberately far beyond the dimensions in the handbook. The scale of that change is a research ambition, not something we expect from a single edit.

The difference between a long pod and a useful harvest

Consider two shrubs. One grows fewer pods, each twice as long as before. The other grows ordinary pods but holds its ripe seeds for another week. Depending on seed number and weight, the second might deliver more protein to the harvest basket. Pod length, by itself, is an unreliable measure of what we want.

We now track the outcome as **recoverable true protein**, rather than as pod length or crude protein percentage. For a single shrub, that means the number of pods produced, seeds per pod, dry mass per seed, the proportion actually harvested, the true-protein fraction of the seeds, and how much survives processing. An additional field-scale model would have to account for plant density, survival, time to maturity, and repeated harvests.

> **Harvested protein = formed seed mass × fraction retained × true-protein fraction × processing recovery**

The distinction between *crude* and *true* protein is important. Converting a plant's total nitrogen into a conventional protein number can count nitrogen that is not part of protein molecules. For food, we would want direct protein measurements and an amino-acid profile, together with tests of the final product.

A soybean with bigger seeds and fewer pods

One useful warning comes from soybean. Researchers comparing variants of *GmCYP78A10* found that an allele associated with larger, heavier seeds was also associated with **fewer pods**. Total seed weight per plant did not change significantly in that study. If we borrow only the attractive half of the finding, we risk designing a plant that looks more productive and is not. [2]

Work in pea gives us a different opening. Developing seeds and the hormones auxin and gibberellin help regulate early growth of the pod wall, including cell division and elongation. That is evidence about a mechanism in pea, not an instruction for making a giant Caragana pod. Still, it tells us where to look for a connection between the seeds and the structure that carries them. [3]

Our first arithmetic model makes this separation explicit. In one *hypothetical* comparison, increasing seed retention from 80% to 95%, with everything else unchanged, increases recovered protein by 18.75%. A sufficiently large increase in pre-harvest damage would cancel that advantage. These are consequences of supplied assumptions, not predicted effects of a genetic edit.

A genome is available; a recipe for editing is not

The genetic side of the investigation has changed quickly. In 2025, Xingyong Cui and colleagues published a chromosome-level reference genome of *C. arborescens*: about 1.216 billion base pairs, assembled into eight pseudochromosomes, with 34,811 predicted protein-coding genes. This is a substantial resource for finding and comparing genes. It is not yet a list of genes that make Caragana pods longer. [4]

A further study published in 2026 assembled genomes from 14 *Caragana* species and built a comparative pan-genome. Among other things, it describes genetic variation associated with responses to dry environments. For our purposes, the opportunity is comparative: relatives can help us distinguish conserved functions from changes particular to one lineage. We have located these genomic resources, but have not yet completed a sequence-based identification of Caragana pod-development genes. [5]

CRISPR is the engineering method we want to investigate, rather than a footnote to conventional breeding. Yet editing a shrub requires more than knowing which DNA sequence to change. The altered cells must give rise to a viable plant, the edit must be confirmed, and its effects must be measured. In 2023, a team established regeneration and stable *Agrobacterium*-mediated transformation in the related species *C. korshinskii*. The authors reported low transformation efficiency and said that they had **not yet tested CRISPR** in that system. A workable *C. arborescens* editing protocol remains an open requirement for us. [6]

One gene name, several very different stories

Pod shattering initially drew us toward soybean's *PDH1* gene. But a legume-wide comparison found that the relevant gene's characteristic genomic neighbourhood is concentrated in the phaseoloid legumes, a group that does not include Caragana. There is even an unrelated soybean gene bearing the same abbreviated name. We therefore set aside any easy claim that the soybean edit can be copied directly into our shrub. [7]

Comparative work in pea and *Medicago ruthenica* instead points to several families involved in pod-wall development and lignin or cellulose accumulation, including AP2/ERF regulators and laccases. These remain **candidate mechanisms**, not identified *C. arborescens* targets. Gene-family names alone cannot establish function. [8]

What the plant keeps for itself

Our original question about protein led into the chemistry of defense. In 1976, researchers isolated two lectins from *C. arborescens* seeds, including a component with strong activity in a laboratory red-blood-cell agglutination assay. Later work identified six soyasaponins in the seeds. Both findings matter to the possibility of food use, but neither paper demonstrates that a particular Caragana preparation is safe or unsafe to eat. That requires separate evidence about dose, bioactivity, processing, and consumption. [9] [10]

Lectins introduce a particularly interesting complication: **lectins are proteins themselves**. Reducing a lectin may improve a food ingredient's suitability, but it cannot automatically increase the amount of total seed protein. A lower concentration of an unwanted protein, a higher concentration of useful proteins, and a larger harvest of safe protein are three distinct possible outcomes. We need assays capable of telling them apart.

Two other shrubs, and a different defense budget

Alongside Caragana we have been examining gorse (*Ulex europaeus*) and Scotch broom (*Cytisus scoparius*), especially their leaves. In a 2021 laboratory study, the reported total protein was about 116 milligrams per gram of dry gorse material and 112 for broom. Their buffer-extractable fractions were much smaller, about 47 and 40 milligrams per gram respectively. A measured protein pool is already a useful observation; it is not yet an edible, safely processed protein ingredient. [11]

These shrubs contain nitrogen-bearing quinolizidine alkaloids, unlike the seed soyasaponins we investigated in Caragana. That difference invites a question: might reducing an alkaloid pool release resources that a plant could use for protein? Our first mass-balance calculations show why this must be treated as a hypothesis. A stock of nitrogen in a defensive compound does not tell us the rate at which that nitrogen could be redirected into seed or leaf proteins.

We also have reasons to keep the defenses. In a common-garden study of gorse, higher concentrations of the alkaloid lupanine were associated with lower infestation by a seed predator and a rust pathogen. Correlation does not establish what would happen after a gene edit, but it cautions against treating defensive chemistry as surplus material waiting to be converted into lunch. [12]

There is an ecological dimension as well. Siberian peashrub has become invasive in some North American habitats. Any proposal to grow or modify it for food would need site-specific containment and ecological assessment, not simply an argument from its hardiness. [13]

What we have actually done

Our present contribution is an organized research and modeling programme. We assembled a registry of perennial-protein research opportunities, with 16 entries in the November 1 working snapshot; these are opportunities for investigation, not sixteen crops judged ready for cultivation. We have linked source-backed evidence about Caragana, gorse, and broom to their existing registry entries and preserved a review copy. The status of the authoritative Mac registry must still be checked before describing those enrichments as fully installed.

We have also executed software. The second version of our perennial-protein processing simulator ran on the Mac, examining **56 hypothetical washing and extraction scenarios** and passing **44 automated tests**. In the ChatGPT sandbox, an initial seed-protein model passed six tests. Its harvest-and-retention successor produced four synthetic scenarios and passed sixteen tests. The software behaves consistently under its declared assumptions. It has not been calibrated against edited Caragana plants or real food-processing trials.

We surveyed other modeling systems, including BioCro for plant growth and VirtualLeaf for tissue development, as possible components for later work. They have not been executed for Caragana here. We have not edited a plant, grown a field plot, measured true seed-protein yield, or demonstrated that the pod-growth and defense hypotheses hold. This is a research programme with a computational notebook, not a digital twin of the shrub.

A study that could help us choose between possibilities

Our next decisive task is unusually concrete: download the matched *C. arborescens* genome annotation and protein sequences, verify that they belong to the intended assembly, and compare plausible pod-development and shattering genes with their relatives. Gene names and annotation keywords will generate leads. Sequence relationships, neighbourhoods, gene trees, and expression in developing pods are needed before naming credible edit targets.

In parallel, plant measurements could establish a baseline that the simulator currently lacks. We would want pod length and shape, pods per plant, seeds per pod, seed mass, shattering behaviour, harvested yield, and true seed-protein composition across genotypes and developmental stages. Seed lectin activity and saponin composition need to be followed through candidate processing steps, with pest and pathogen responses measured separately. Even a well-documented negative result could eliminate a misleading pathway before years of shrub growth and editing work.

The question that now interests us is more demanding than our first picture of a spectacular pod. **Could a perennial legume reliably provide more recoverable, usable protein while remaining a healthy plant and an acceptable neighbour?** We have several possible routes to investigate. None has yet earned the status of a demonstrated answer.

Sources and methodological notes

1. Dietz, D. R., Slabaugh, P. E., & Bonner, F. T. (2008). "*Caragana arborescens* Lam.: Siberian peashrub." In *The Woody Plant Seed Manual*, USDA Agricultural Handbook 727, pp. 321–323. research.fs.usda.gov/treesearch/32630
2. Wang, X., et al. (2015). "Evolution and association analysis of *GmCYP78A10* gene with seed size/weight and pod number in soybean." *Molecular Biology Reports* 42, 489–496. doi:10.1007/s11033-014-3792-3
3. Ozga, J. A., van Huizen, R., & Reinecke, D. M. (2002). "Hormone and seed-specific regulation of pea fruit growth." *Plant Physiology* 128, 1379–1389. doi:10.1104/pp.010800
4. Cui, X., et al. (2025). "Ancient hybridization drives arid adaptation and species diversification in *Caragana* (Fabaceae)." *New Phytologist* 247, 2454–2472. doi:10.1111/nph.70360
5. Song, J., et al. (2026). "Pan-genome analysis of *Caragana* reveals genomic mechanisms of drought adaptation of leguminous shrubs to arid environments." *Genome Biology* 27, 331. doi:10.1186/s13059-026-04230-8
6. Liu, B., et al. (2023). "In vitro regeneration and *Agrobacterium*-mediated genetic transformation of *Caragana korshinskii*." *Forestry Research* 3, 14. doi:10.48130/FR-2023-0014
7. Marsh, J. I., et al. (2023). "Legume-wide comparative analysis of pod shatter locus *PDH1* reveals phaseoloid specificity, high cowpea expression, and stress responsive genomic context." *The Plant Journal* 115, 68–80. doi:10.1111/tpj.16209
8. Yong, X., et al. (2024). "Paralogous Gene Recruitment in Multiple Families Constitutes Genetic Architecture and Robustness of Pod Dehiscence in Legumes." *Genome Biology and Evolution*. doi:10.1093/gbe/evae267
9. Bloch, R., Jenkins, J., Roth, J., & Burger, M. M. (1976). "Purification and characterization of two lectins from *Caragana arborescens* seeds." *Journal of Biological Chemistry* 251, 5929–5935. doi:10.1016/S0021-9258(17)33041-7
10. Taylor, W. G., Sutherland, D. H., Richards, K. W., & Zhang, H. (2015). "Oleanane Triterpenoid Saponins of *Caragana arborescens* and Their Quantitative Determination." *Industrial Crops and Products* 77, 74–80. doi:10.1016/j.indcrop.2015.08.034
11. Iyer, A., Bestwick, C. S., Duncan, S. H., & Russell, W. R. (2021). "Invasive Plants Are a Valuable Alternate Protein Source and Can Contribute to Meeting Climate Change Targets." *Frontiers in Sustainable Food Systems* 5:575056. doi:10.3389/fsufs.2021.575056
12. Hornoy, B., et al. (2012). "Alkaloid concentration of the invasive plant species *Ulex europaeus* in relation to geographic origin and herbivory." *Naturwissenschaften*. doi:10.1007/s00114-012-0970-9
13. Wisconsin Department of Natural Resources. "Siberian peashrub (*Caragana arborescens*)." Invasive Species Program. dnr.wisconsin.gov/topic/Invasives/fact/SiberianPeashrub

Method note. Published measurements and comparative genetic findings are attributed to their original studies. The 80%-to-95% seed-retention comparison is arithmetic on hypothetical inputs. Mac and sandbox tests verify execution and software behavior, not biological accuracy. Our 16 registry entries are research opportunities rather than validated edible crops. Bibliographic identities and the main cited findings were checked against scientific and institutional records on 9 November 2026. No *Caragana arborescens* edit, field trial, edible true-protein assay, or validated gene-to-pod prediction has been performed by Phasespace as part of this project.