11-20-2015

Discovery and Metabolic Engineering of Steroid Alkaloid Biosynthetic Genes from (*Veratrum californicum*)

Megan M. Augustin
University of Missouri-St. Louis, megan.m.augustin@gmail.com

Discovery and Metabolic Engineering of Steroid Alkaloid Biosynthetic Genes from *Veratrum californicum*

By

Megan M. Augustin

B.S. Biology, University of Missouri-St. Louis, 2007

A.A. General Transfer Studies, St. Louis Community College-Florissant Valley, 2004

A Thesis

Submitted to The Graduate School of the

University of Missouri-St. Louis
in partial fulfillment of the requirements for the degree

Master of Science

In

Biology
With an emphasis in Molecular and Cellular Biology

December 2015

Advisory Committee

Wendy Olivas, Ph.D.
Chairperson

Bethany Zolman, Ph.D.

Toni M. Kutchan, Ph.D.

Abstract

The steroid alkaloid cyclopamine has shown much promise as a treatment for cancers in which aberrant hedgehog signaling plays a role. The compound, originally discovered due to its teratogenic effects in sheep, binds to the hedgehog signaling receptor Smoothend and prevents downstream activation. As this pathway is primarily active during embryonic development, overactivation later in life can lead to tumor formation and proliferation. Several studies have shown that cyclopamine can inhibit and even reverse tumor growth, but limited supply will prevent widespread use upon FDA approval of it or any of its semi-synthetic analogs. As high value plant medicinal compounds are in demand, producing them at industrially feasible levels is not always possible. Many plant natural products are originally produced in low quantities and often from species that are not amendable to cultivation. Production of these compounds in a heterologous system is ideal, but first the underlying genes that encode biosynthetic enzymes must be discovered. Cyclopamine is produced in one of these non-ideal sources, the slow growing *Veratrum californicum*, a species that has not been successfully cultivated. Our goal was to discover the underlying genes in the biosynthetic pathway to cyclopamine and express them in the emerging heterologous production system *Camelina sativa*. Herein, the first four genes in the biosynthetic steroid alkaloid pathway to verazine, a hypothesized intermediate of cyclopamine, have been discovered by correlating gene expression and alkaloid accumulation. In addition, the genes have been successfully introduced into *C. sativa* under the control of seed specific promoters for the production of *V. californicum* secondary metabolites in seed. Future work requires discovery of the remaining cyclopamine biosynthetic genes and metabolic fine tuning for increased metabolite yield, but the groundwork has been set for future work using *C. sativa* as a production system for high value, medicinally significant, plant natural products.

Acknowledgements

I would like to sincerely thank Dr. Toni Kutchan for providing me with the resources, expertise, and guidance required to perform the work herein as all research was conducted in her laboratory. I would also like to thank my advisor, Dr. Wendy Olivas, for her invaluable academic guidance required throughout my graduate career. In addition, I want to thank Dr. Bethany Zolman for her time and efforts required as a member of my graduate committee. I would also like to thank Infinity Pharmaceuticals, Inc. for funding the initial work for this project and Dr. Davin A. Mann for the collection of wild *Veratrum californicum* used in this study. In addition, I would also like to thank the members of the Kutchan laboratory for their advice and opinions at the bench. Specifically, I want to thank my husband, Jörg M. Augustin, for his exceptional advice, suggestions, and support. Finally, I would like to thank the Donald Danforth Plant Science Center and the Proteomics and Mass Spectrometry Facility, as the work could not have been completed without the important resources they provided.

Contents

Chapter 1: History and introduction of cyclopamine

1.1 History of *Veratrum californicum* and cyclopamine

Peculiar birth defects ranging from pronounced lower jaw to full blown cyclopia were reported to occur in lambs born in southwestern and south central Idaho from the early 1900-1950's (Figure 1)**.** The incidence of congenital anomalies occurred in up to 25% of offspring in some sheep herds (Keeler 1978a). Frequencies were not necessarily accurate as some sheep herders failed to report incidences due to fear of gaining a negative reputation of owning poor genetic stock. Appropriate breeding practices, however, made genetic causes unlikely (Keeler 1978a). In addition, reports of these birth defects appeared to correlate with grazing location as opposed to specific breeding lines. Ranchers finally contacted the Poison Plant Research Laboratory (PPRL) of the United States Department of Agriculture in 1954 for help in determining the cause behind the instances (Lee *et al.* 2014).

Figure 1. Head of a lamb whose mother ingested *Veratrum californicum.* Adapted from Lee et al. 2014.

The PPRL, located in Logan, Utah, helped initially rule out genetic causes through breeding experiments by mating ewes known to have birthed deformed lambs with their own rams (Binns *et al.* 1959).These results indicated what was previously expected; the origin of the birth defects was likely caused by foraging. Continued investigation led researchers to conduct feeding trials based upon a plant common to all ranges where the deformities were reported. The malefactor was officially determined to be a member of the Melanthiaceae/Liliaceae family (debated phylogenetic placement), *Veratrum californicum* (false hellebore, corn lily), a native plant frequently found in mountainous regions of the Northwestern United States (Figure 2) (Binns *et al.* 1963, Heretsch *et al.* 2010). A range in birth defects from mild to severe was observed upon ingestion of *V. californicum* by pregnant ewes, including holoprosencephaly (lack of division in forebrain), single and double globe cyclopia containing only a single optic nerve, and a proboscis emerging from the forehead. In addition,

Figure 2. Photograph of *Veratrum californicum* courtesy of Jerry Friedman.

the upper palate was often missing in conjunction with a pronounced lower jaw. The most severe cases resulted in aborted lambs (Binns *et al.* 1962, Keeler 1978a). Continued investigation indicated that the birth defects correlated with the time of *V. californicum* ingestion. The highest incidence of cyclops lambs occurred when *V. californicum* was consumed at day 14 of gestation, the time when the primitive streak develops in sheep (Binns *et al.* 1965). Consumption of *V. californicum* at subsequent periods of gestation resulted in malformations of the trachea, gut, and limbs (Lee *et al.* 2014). Now that the causative species and knowledge concerning the time of insult for *V. californicum* exposure was revealed, ranchers could minimize incidences of abnormal births by directing animal foraging. These practices have successfully kept re-occurrences of birth defects due to *V. californicum* poisoning low (Keeler 1978b).

Even after the discovery of *V. californicum* as the cause of congenital abnormalities and implementation of effective grazing management, researchers at the PPRL continued to search for the precise teratogenic agent. Previous studies indicating the inhibitory effects that certain steroidal alkaloids had on cell division hinted to researchers that this class of metabolites might be of interest (Keeler 1978a). Initial investigations were conducted by testing fractions of different *V. californicum* tissue extracts for teratogenicity. The compounds causing the abnormalities were shown to be extractable by benzene/ammonia followed by ethanol (Keeler and Binns 1966a, Keeler and Binns 1966b). Subsequently, steroidal alkaloids, both the glycosides and parent alkamines from *V. californicum,* were extracted, purified, and tested for teratogenicity in sheep. In addition, structurally related compounds were tested. Only three steroidal alkaloids, obtained directly from *V. californicum,* were found to cause the teratogenic effect of cyclopia in lambs. These included cyclopamine **(1)** (the most potent), cycloposine **(2)** (glycoside of cyclopamine), and jervine **(3)** (Figure 3)(Keeler and Binns 1968). Additional testing with other structurally related steroid based compounds including alkaloids, sapogenins, and hormones were conducted in order to elucidate the exact structural component that conferred teratogenicity. The results were not 100% conclusive, but point to the presence of the furan ring and its respective ether bridge (Keeler 1970a). Further analysis of the teratogenicity of the related steroidal alkaloid solasodine **(4)** provided evidence that the nitrogen in the F-ring was required, and its projection in the alpha position may be essential. The basic nature of the nitrogen is higher in cyclopamine, which may account for its greater teratogenic effects compared to solasodine (Keeler *et al.* 1976). One early hypothesis concerning the teratogenicity of cyclopamine included conversion to an active form in the gut of ruminant animals, but non-

ruminant animals such rabbits, hamsters, rats, and mice were also susceptible to cyclopamine poisoning (Binns *et al.* 1972, Keeler 1970b, Keeler 1975). In addition, direct application of cyclopamine to developing embryos also caused teratogenic effects (Bryden and Keeler 1973, Bryden *et al.* 1973).

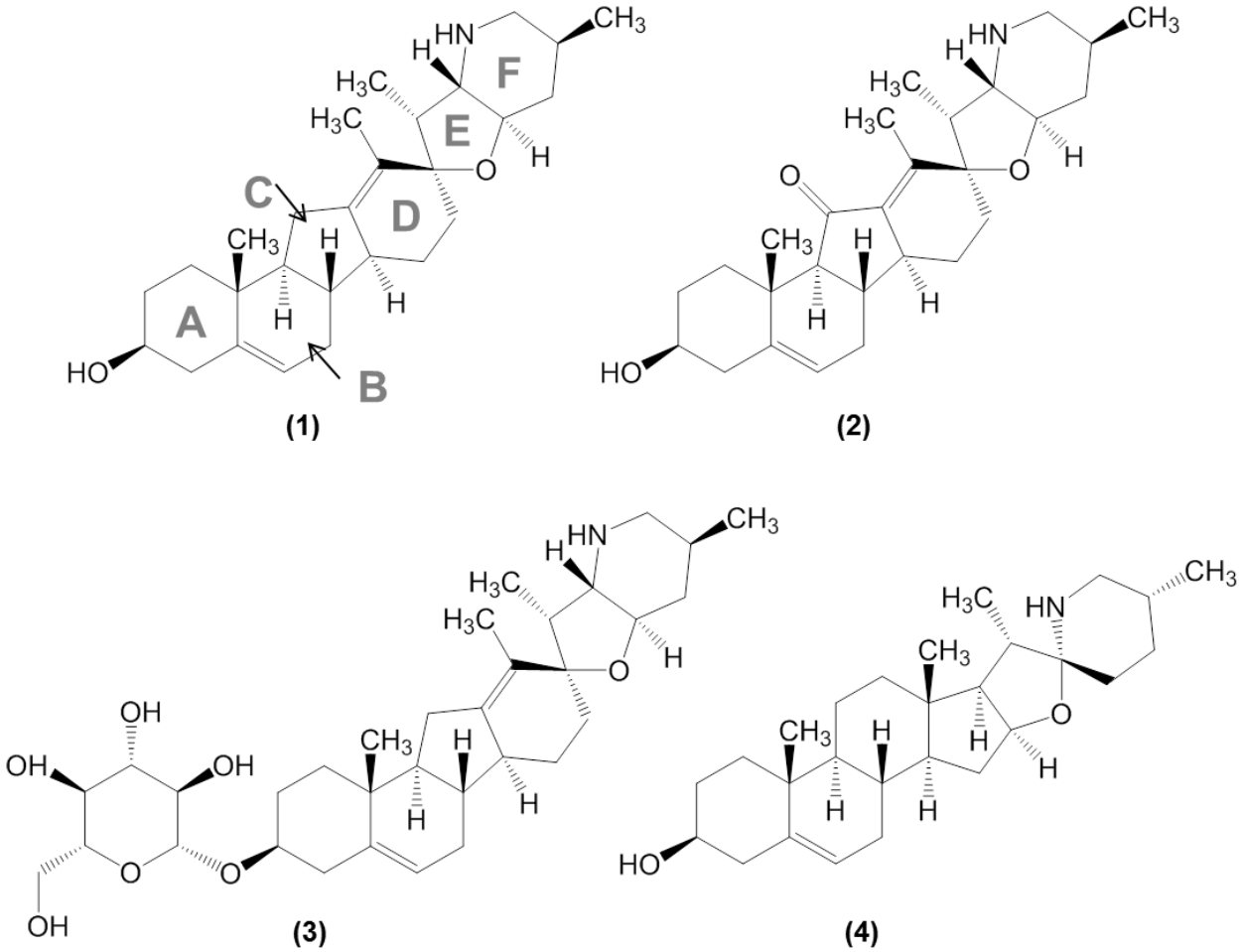

Figure 3. Structures of three teratogenic steroid alkaloids from *Veratrum californicum* including cyclopamine (1), jervine (2), and cycloposine (3) and the structurally related teratogenic solanum alkaloid solasodine (4).

Structural elucidation of cyclopamine was discovered first in *Veratrum grandiflorum* by a research group in Japan (Masamune *et al.* 1964). It was unknown at the time that this compound was identical to the compound causing cyclopic teratogenicity in lambs in the United States. The structure was similar to the known structure jervine, only missing the ketone moiety on the C-11 position, and therefore called 11-deoxojervine. The steroid alkaloid in *V. californicum* was later confirmed to be identical to that in *V. grandiflorum* by the researchers at the PPRL (Keeler 1969). Cyclopamine contains six rings (A-F) and possesses a C-*nor*-D-*homo*

ring structure (Figure 3)**.** The F-ring is a piperidine structure containing a secondary basic nitrogen. The D-ring is connected to the tetrahydrofuran E ring by a spiro connection (Heretsch *et al.* 2010). More information on steroid alkaloids and their structures is described below.

1.2 Diversity of steroid alkaloids

A few classes of plants and animals synthesize a wide range of structurally diverse and often bioactive steroid alkaloids. The plant families containing the majority of steroid alkaloids (in bulk and diversity) include Apocynaceae, Buxaceae, Liliaceae, and Solanaceae. Steroid alkaloids have also been discovered in a few vertebrate species and marine animals. Marine invertebrates including *Zoanthid*, *Cephalodiscus*, and *Ritterella* in addition to the amphibians *Salamandra* (a genus of six salamanders), *Bufo* (true toads) and *Phyllobates* (a genus of poison dart frogs) also contain steroid alkaloids. Many have been shown to have significant biological activity including therapeutic benefits as well as microbial/fungal growth inhibition. The origin of most steroid alkaloids derives from the triterpene skeletons cycloartenol, lanosterol, and cholesterol, followed by a numerous combination of chemical alterations. Due to the steroid or triterpene origin, they are often referred to as steroidal amines as opposed to true alkaloids that derive from amino acids (Cordell 1998).

Members of the Apocynaceae family (including *Holarrhena, Paravallaris, Funtumia, Kibatalia*, and *Malouetia)* generally produce steroid alkaloids of two basic types, the conanine-type **(5)** and pregnane-type **(6)** (Figure 4). The pregnane-type is medicinally useful as alkaloids of this class may be converted to steroid hormones. Other biological activities from Apocynaceae alkaloids include activity against amoebiasis conferred by the conanine-type steroid alkaloid conessine and hepatocarcinogenesis conferred by the pregnane-type steroid alkaloid irehdiamine-A (Cordell 1998).

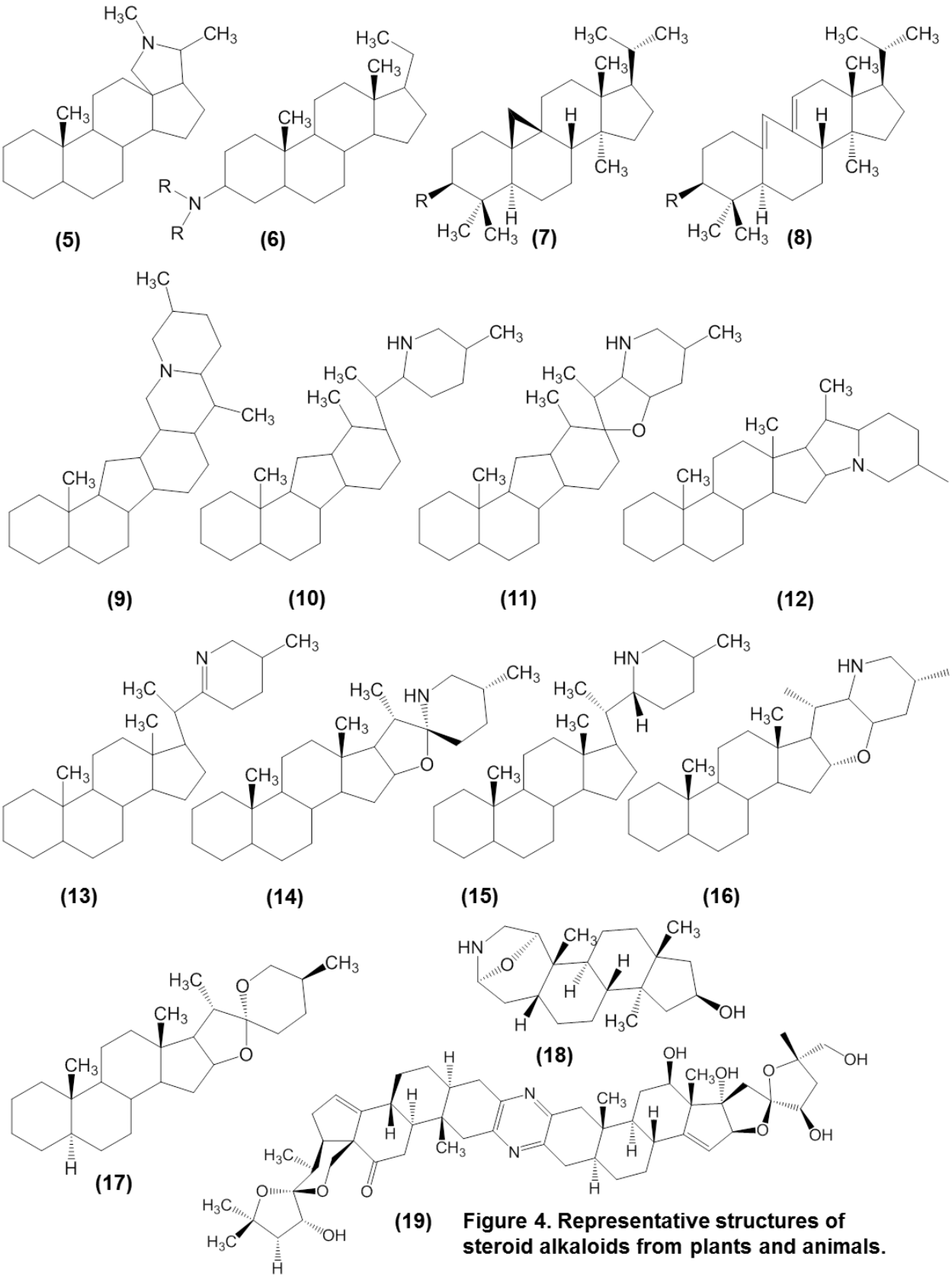

Figure 4. Representative structures of steroid alkaloids from plants and animals.

Several genera of the Buxaceae are known to contain steroid alkaloids including *Buxus*, *Sarcococca*, and *Pachysandra*. There are two main skeletal structures found in this family (Figure 4) one including a characteristic cyclopropane ring **(7)** and one that does not **(8)**. Structures of these classes often contain methyl groups at one or both C-4 and C-14 positions, making them intermediates between cholesterol and lanosterol type steroids. Some genera also contain the pregnane-type steroid alkaloids. Extracts of boxwood (*Buxus sempervirens*) contain the steroid alkaloids cyclobuxine-D and buxamine, and have exhibited medicinal effects against several ailments including malaria, syphilis, cancer, and tuberculosis (Cordell 1998).

Genera of the Liliaceae known to contain steroid alkaloids include *Veratrum, Fritillaria, Petilium, Zygadenus,* and *Korolkowia.* Three general classes of steroid alkaloids are found in the Liliaceae including the cerveratrum-type, jerveratrum-type, and the solanidine-type. These classes can be further divided into 5 distinct groups, the cevanine-type **(9)** (cerveratrum-type), the veratramine-type **(10)** (distinct from jerveratrum-type as it lacks the ether bridge of jervine), the jervine-type **(11)**, the solanidine-type **(12)**, and the verazine-type **(13)** (distinct from the solanidine-type as it contains only 5 rings) (Figure 4) (Li *et al.* 2006). Compounds containing the ring structure similar to cholesterol (6-6-6-5) are considered steroid alkaloids (the solanidine and verazine-types) and compounds containing a C-*nor*-D-*homo* ring system (6-6-5-6) are considered isosteroid alkaloids (cevanine, veratramine, and jervine types). The designation C-*nor*-D-*homo* refers to the C-ring having one less carbon ('*nor*') and the D-ring having an additional carbon ('*homo*'). Cyclopamine falls into the jervine-type as it possesses a hexacyclic skeletal system with a C-*nor*-D-*homo* ring structure in addition to an ether bridge between the D and F-rings, forming the tetrohydrofuran E-ring. In addition, a spiro connection is found between D and E-ring at the C-17 position. Cevanine-type alkaloids are highly hydroxylated with a distinctive C-4, C-9 linkage between the A and B-rings. Alkaloids of the veratramine type are missing the ether bridge and therefore lack a 6^{th} ring in conjunction with a characteristic aromatic D-ring (Cordell 1998). Plants of these genera have been used in traditional medicines to treat ailments including toothache, sore throat, venereal diseases, wounds, epilepsy, and cardiac disease. In addition, crushed *Veratrum viride* rhizome was applied to snakebites and the European *Veratrum* species *V. album* was known for its emetic properties. Treatment of hypertension with the steroid alkaloid protoveratrine is also known (Chandler and McDougal 2014).

Five types of steroid alkaloids found in the Solanaceae (including the genera *Solanum* and *Lycopersicon*) are the solanidine-type **(12)**, spirosolane-type **(14)**, solacongestidine-type **(15)**

solanocapsine-type **(16)**, and jurbidine-type **(17)** (Figure 4). The solacongestidine-type is essentially the same as the verazine-type, but different sources have slightly different classifications, so both are relevant terms. Medicinal value from Solanaceae alkaloids is marked as they can serve as precursors to steroidal hormones. Antimicrobial activity has also been detected from several compounds (Cordell 1998). In addition, α-tomatine, a glycosylated steroid alkaloid from *Lycopersicon esculentum*, has exhibited antineoplastic activity in prostate cancer cells (Lee *et al.* 2011, Lee *et al.* 2013a, Lee *et al.* 2013b).

The distinct types of steroid alkaloids are not necessarily unique to each genus as several alkaloids of the same type can be found among the different families. For example, the solanidine and verazine-type can be found in the Liliaceae and the Solanaceae. Also, the pregnane-type is found in both the Apocynaceae and Buxaceae. However, the cevanine, veratramine, and jervine-types that contain the characteristic C-*nor*-D-*homo* ring structure are, so far, found exclusively in the Liliaceae (Cordell 1998).

Steroid alkaloids from the animal kingdom can also differ markedly from those found in the plant kingdom. For example, the structurally distinct samandarin **(18)** is a characteristic steroid alkaloid from *Salamandra* species containing an expanded A ring (Figure 4). Also of note, the nitrogen atom was found to derive from glutamine, contradictory to the predicted source for nitrogen for steroid alkaloids in plants (discussed below). Also quite distinct from the steroid alkaloids in plants are those found in marine organisms. These include the novel dimeric steroid backbones like that found in cephalostatin **(19)** (Figure 4) (Cordell 1998).

1.3 Hedgehog signaling and cyclopamine

1.3.1 Discovery of hedgehog pathway

The teratogenic effect of cyclopamine was eventually discovered in the late 1990s. Inhibition of Hedgehog (Hh) signaling in developing embryos resulted in the craniofacial birth defects observed (Incardona *et al.* 1998). The Hh pathway was originally discovered in Heidelberg at the European Molecular Biology Laboratory in the late 1970s by Nüsslein-Volhard and Wieschaus during mutational studies and their effects on embryogenesis in *Drosophila melanogaster*. In *D. melanogaster*, this signaling pathway is primarily responsible for segmentation and development of the body plan. One of the fifty mutants in the screen resulted in larva covered in epidermal spikes. This mutant led to the naming, cloning, and identification of the first Hh gene.

The Hh signaling pathway is highly conserved from flies to mammals as it plays a predominate role in embryonic development. It is responsible for many developmental processes in mammals including formation of the limbs, nervous system, teeth, lungs, intestine, cartilage, and sperm and is heavily involved in bilateral symmetry. Three well conserved homologs to the *D. melanogaster* Hh gene exist in mammals; Sonic Hedgehog (Shh), Desert Hedgehog (Dhh), and Indian Hedgehog (Ihh). Each are involved in unique processes, with Shh being the most abundant ligand, having the broadest expression and influence and being the most well studied (Heretsch *et al.* 2010, Lee *et al.* 2014). The hedgehog signaling pathway is primarily silenced in adults, but reactivation can occur for regeneration and injury repair. Aberrant overactivation can lead to uncontrolled cell growth and oncogenesis (McMillan and Matsui 2012). Cyclopamine inhibition of the Hh pathway has been shown to downregulate overactivation, stopping and even reversing tumorigenesis caused by abnormal signaling (Taipale *et al.* 2000).

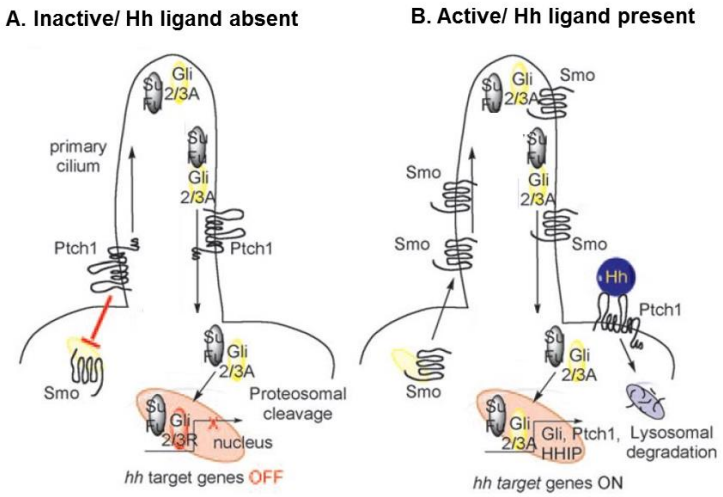

Figure 5. Canonical Hedgehog (Hh) signaling pathway. A) Pathway inactivated in absence of Hh ligand. PTCH suppresses SMO and prevents movement into the primary cilium and subsequent activation of Hh response genes. In absence of SMO signaling, SUFU and other factors process the transcription factors Gli2/3 into repressor form. B) Pathway activation by Hh ligand initiates degradation of PTCH and accumulation of SMO in the primary cilium. SMO prevents processing of Gli transcription factors to repressor form thereby allowing the activation of Hh response genes. Figure adapted from Heretsch et al. 2010.

1.3.2 Canonical hedgehog signaling

Canonical hedgehog signaling is activated by the presence of the Hh ligands (Shh, Ihh, and Dhh in mammals). Initially, the signal sequence is removed from the nascent Hh peptide to yield a ~45 kD *Hh* precursor protein. Then, activation of the ligand begins when the protein undergoes autocatalytic-cleavage initiated by the C-terminal domain leaving a 20 kD N-terminal domain and a 25 kD C-terminal domain. In addition, a cholesterol moiety is added to the C-terminal end of the N-terminal fragment. The C-terminal fragment functions solely in processing while the N-terminal fragment serves as the signaling molecule. The cholesterol modification results in association of the Hh signaling peptide with the membrane, allowing the addition of a palmitoyl moiety to the N-terminal end by Skinny Hedgehog (Ski), an acetyl transferase protein. The final result is an active signaling molecule (Porter *et al.* 1996, Varjosalo and Taipale 2007). In the

absence of active Hh, the 12-transmembrane receptor Patched (PTCH) represses the Hh pathway by preventing the 7-transmembrane G-protein coupled receptor-like Smoothened (SMO) from entering the primary cilium, a specialized cellular projection involved in signaling. Upon activation by Hh signaling proteins, PTCH leaves the primary cilium and is degraded by endosomes, resulting in an accumulation of SMO in the primary cilium and activation of the pathway (Corbit *et al.* 2005, Peluso *et al.* 2014, Rohatgi and Scott 2007). Ligand binding to one of the co-receptors of PTCH (CDO, BOC, or GAS1) is also required for activation (Izzi *et al.* 2011). Once SMO is active, it prevents the proteolytic processing of Gli2/3 (zinc finger transcription factors) to their repressor forms by inhibiting the negative regulator Suppresser of Fused (SUFU), Casein Kinase 1(CK1) Protein Kinase A (PKA), and Glycogen Synthase Kinase 3β (GSK3β) (Cochrane *et al.* 2015, Hui and Angers 2011). In the absence of signal, Gli2/3 are processed into truncated forms that serve as repressors of the hedgehog target genes. Therefore, active forms of Gli2/3 induce expression of hedgehog response genes. In addition, a positive feedback loop is initiated as active Gli2/3 induce expression of Gli1, a protein involved in amplification of the signal. PTCH expression is also induced by hedgehog pathway activation, serving to downregulate the signal, thereby playing an opposing role to Gli1 (Figure 5) (Hui and Angers 2011, McMillan and Matsui 2012, Niewiadomski *et al.* 2014).

1.3.3 Mechanism of hedgehog inhibition by cyclopamine

The exact mechanism of Hh signaling inhibition by cyclopamine was discovered by the research group of Philip Beachy at the Johns Hopkins University School of Medicine with the help of collaborators at the Howard Hughes Medical Institute. Beachy's group was studying hedgehog signaling in mammals and generated a cyclops mouse upon mutation of the Shh gene. The connection was then made between the outbreaks of cyclops lambs in the 1950's and Hh signaling. Searching for an efficient way to control Hh signaling, his group turned to cyclopamine, the causative agent of the teratogenic effects leading to cyclopia (Heretsch *et al.* 2010, Keeler and Binns 1968). Beachy's group first hypothesized that cyclopamine inhibited cholesterol biosynthesis, but it was shown that signal transduction was perturbed and the synthesis of cholesterol was not involved (Cooper *et al.* 1998, Incardona *et al.* 1998). In the end, their research showed that hedgehog signaling was inhibited when cyclopamine bound directly to SMO and prevented downstream signaling, regardless of Hh ligands(Chen *et al.* 2002). More recent work has shown that native oxysterols bind to the extracellular domain of SMO and are required for its ability to send downstream signals. This binding site is distinct from that bound by cyclopamine, however, as cyclopamine binds to the heptahelical bundle, a transmembrane region of SMO (Nachtergaele *et al.* 2012, Nedelcu *et al.* 2013, Sharpe *et al.* 2015).

1.3.4 Abnormal hedgehog signaling and cancer

Indication of the relationship between the development of cancer and abnormal signaling of the Hh pathway originated in patients with the autosomal dominant Gorlin syndrome (Basal-cell nevus syndrome). Patients suffering from Gorlin syndrome have skeletal abnormalities, basal cell carcinomas, cysts, and cleft lip/palate in addition to other neoplasms and birth defects (Gorlin 1995). It was discovered that this syndrome could be attributed to germline mutations in PTCH, and therefore abnormal Hh signaling (Hahn *et al.* 1996, Johnson *et al.* 1996). Mutations in PTCH have also been identified in up to 90% of sporadic basal cell carcinomas (Epstein 2008). In addition to Gorlin syndrome and basal cell carcinomas, development of medulloblastoma has also been associated with reduced function of PTCH (Corcoran and Scott 2001, Wechsler-Reya and Scott 2001).

Aberrant Hh signaling has been subsequently shown to be involved in the development and progression of numerous cancers through several mechanisms. One mechanism is through ligand-*independent* constitutive activation. In this mechanism, overactivation of the Hh pathway occurs regardless of Hh signaling protein. Mutations in PTCH (e.g. Gorlin syndrome and medulloblastoma), SMO, or SUFU can fall into this category. Other mechanisms include autocrine ligand-*dependent* (cells respond to over production of its own Hh signal) and paracrine ligand-*dependent* (cells respond to overproduction of Hh signals from nearby cells). More evidence supports the involvement of paracrine signaling in cancer growth, highlighting the importance of the nearby stromal cells (connective tissue) and the surrounding microenvironment. Some research, however, has shown the possibility of autocrine signaling playing a role in acute lymphocytic leukemia (Lin and Matsui 2012, Lin *et al.* 2010, Yauch *et al.* 2008).

Ligand-dependent overactivation of the Hh pathway has been found to play a role in more cancer types than ligand-independent overactivation. Production of Hh paracrine factors from stromal cells was discovered to assist in the survival of B-cell malignancies, demonstrating the importance of the microenvironment in tumor production (Dierks *et al.* 2007). Overexpression of Shh was observed in 70% of pancreata samples from patients with pancreatic carcinoma indicating a loss of Hh regulation. This misregulation was not attributed to mutations in PTCH, SMO, or SUFU and is therefore ligand-dependent. Expression of PTCH, not detected in normal pancreata cells, was also seen in abnormal epithelium and the surrounding mesenchymal cells (Thayer *et al.* 2003). Ligand-dependent overactivation of Hh pathway was also observed in some digestive tract cancers, small cell lung cancers, myeloid leukemia, multiple myeloma,

colon cancer, melanomas, and prostate cancers (Berman *et al.* 2003, Karhadkar *et al.* 2004, Peacock *et al.* 2007, Stecca *et al.* 2007, Varnat *et al.* 2009, Watkins *et al.* 2003, Zhao *et al.* 2009). Aberrant Hedgehog signaling was also observed in malignant pleural mesothelioma (Shi *et al.* 2012).

Abnormal Hedgehog signaling is also involved in the proliferation and maintenance of cancer stem cells. Using multiple myeloma as a model, researchers discovered the requirement of Hh signaling in proliferation of tumor stem cells. The presence of Hh was also necessary to keep the stem cells in an undifferentiated state (Peacock *et al.* 2007). The role of Hh signaling has also been established in colon cancer stem cells and myeloid leukemia stem cells (Batsaikhan *et al.* 2014, Varnat *et al.* 2009, Zhao *et al.* 2009). Cancer stem cells are attributed to the reoccurrence of many cancers and are a particularly important target of cancer treatment.

1.3.5 Hedgehog signaling inhibition as cancer treatment

Inhibition of hedgehog signaling in many cancers where abnormal regulation is implicated has shown great promise. Treatment of medulloblastoma cell cultures with cyclopamine showed specific inhibition and reduction of tumor cell lines by up to 80% while fibroblast cells were unaffected (Berman *et al.* 2002). In another study, cell death and decreased tumor size was observed in a medulloblastoma mouse model by preventing hedgehog signaling (Romer *et al.* 2004). In prostate cancer, overexpression of PTCH was observed up to 400-fold the normal levels in cancer cells as compared to benign epithelial cells. Treatment with cyclopamine reversed these effects and inhibited growth in three out of four cell lines tested. In the same study, cyclopamine treatment of prostate tumors in mouse models showed suppression and regression. Discovery that Hh pathway activation is required in prostate cancer stem cells in conjunction with the observed tumor regression indicates the cancer stem cells were also effectively treated with cyclopamine (Karhadkar *et al.* 2004). Increased expression of Hh pathway target genes was also observed in colon cancer cells, with a marked increase in metastatic tumors. Treatment of colon cancer with cyclopamine inhibited hedgehog signaling, caused reduced cell proliferation, increase apoptosis, and even eliminated tumors in mouse models for up to one year when treated for 20 days. Treatment of melanoma cultures with cyclopamine for 20 days also caused cancer cell death and prevented tumor recovery. Shorter treatment periods in the colon cancer study initially eliminated tumors but regrowth was observed, highlighting the necessity of Hh signaling for tumor reoccurrence (Stecca *et al.* 2007, Varnat *et al.* 2009). A reduction in the expression of stem cell specific genes was also seen when colon cancer cell cultures were treated with cyclopamine, indicating a possible role in

treatment for colon cancer stem cells (Batsaikhan *et al.* 2014). As for pancreatic cancer, cyclopamine treatment was found effective in 50% of cancer cell lines tested. Cyclopamine induced apoptosis, reduced cell density, and tumor mass was reduced up to 60% in a mouse model (Thayer *et al.* 2003). Treatment of mouse models with myeloid leukemia showed a 14-fold reduction in cancer stem cells. In addition, 60% of cyclopamine-treated mice were alive after 7 weeks while the control mice did not survive over 4 weeks. It was also noted that earlier treatment with cyclopamine was more effective (Zhao *et al.* 2009) In research with multiple myeloma, treatment with cyclopamine inhibited overactivation of Hh pathway and diminished clonal capacity of cancer stem cells (Peacock *et al.* 2007). Another study showed complete abolition of lymphoma cancer cells grown on a layer of non-cancerous stromal cells within 48 hours while the stromal cells were unaffected (Dierks *et al.* 2007). Targeted inhibition of the Hh pathway in cancer cells vs. normal cells by cyclopamine and the marked effects on cancer stem cells shows positive attributes for future cancer treatment.

Further elaboration on the importance of the microenvironment in tumor growth and its potential as a therapeutic target was highlighted in two contradictory studies involving the effects of cyclopamine / IPI-926 and pancreatic cancer. In one study, the treatment of pancreatic tumors with cyclopamine showed a significant reduction of 42% in microvascular density. Angiogenesis (growth of new blood vessels) was inhibited and the remaining tumor vasculature was narrow and fragmented (Nakamura *et al.* 2010). A study published one year earlier, however, found that treatment with the cyclopamine analog IPI-926 increased the vascular density and increased the delivery of the chemotherapeutic agent gemcitabine (Olive *et al.* 2009). This could possibly be due to the difference between cyclopamine and its analog, but the actual effects on cyclopamine and growth of vasculature are still debated.

Several structural variations of cyclopamine have been designed for better solubility and stability in order to increase its effectiveness as a cancer treatment. The double bond between C-12 and C-13 creates an allylic oxygen in the furan ring, thus acid treatment (as would be encountered in the stomach) will cause a break in the ether bridge, followed by aromatization of the D ring resulting in toxic veratramine. KAAD-cyclopamine **(20)** (Figure 6), an analog with an attachment to the nitrogen, was shown to have 10-20 times the potency of cyclopamine (Taipale *et al.* 2000). Carbohydrate derivatives of cyclopamine prepared with glycosyl azides in a manner called “Click Chemistry” (easy chemical modifications under mild conditions) was utilized in an attempt to increase solubility. Only one compound of the several tested came close to the activity of cyclopamine (Zhang *et al.* 2008). In attempts to diminish the potential toxic side

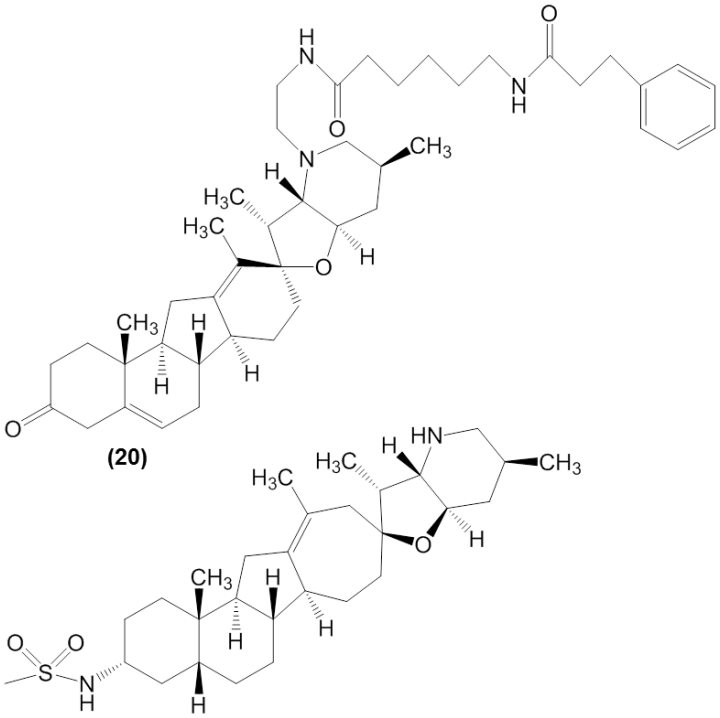

Figure 6. Structural analogs of cyclopamine including KAAD-cyclopamine (20) and IPI-926 (21).

effects of cyclopamine on non-cancer cells, some groups have developed cyclopamine prodrugs. In one study, cyclopamine was conjugated to a protein cleavable by prostate-specific antigen (PSA), a serine protease that is highly expressed in prostate cells; cyclopamine is only activated in these cells. A 7-fold enhancement of the cyclopamine derivative was detected in the presence of PSA as opposed to controls without PSA (Kumar *et al.* 2008). Another group designed a cyclopamine analog bearing a glycosylated poly(ethylene glycol) side chain that is cleaved by β-glucuronidase, an enzyme known to be highly concentrated in necrotic areas of many tumors (Renoux *et al.* 2011). None of these analogs have, so far, progressed passed the research laboratory.

Another promising cyclopamine derivative is IPI-926 **(21)** (Figure 6), advanced by Infinity Pharmaceuticals from a batch of semi-synthetic derivatives including the seven-membered-ring containing analog IPI-269609. This molecule showed increased solubility and stability with positive results in lab settings. Complete regression of a medulloblastoma model was observed in tumor bearing mice treated orally IPI-926 as opposed to cyclopamine treatments which showed little effect under these conditions (Tremblay *et al.* 2009). Initial clinical trials also showed positive results for patients with basal cell carcinomas (Jimeno *et al.* 2013). Less promising results were obtained, however, in Phase 1b/II clinical trials for IPI-926 treatment of pancreatic cancer in conjunction with Gemcitabine and Phase II clinical trials with myelofibrosis. Both studies were terminated early as patients either had a lower median survival rate than the controls or showed no improvement (Lee *et al.* 2014). Another Phase I clinical trial is currently recruiting participants for IPI-926 in conjunction the chemotherapeutic drug FOLFIRINOX (Ko 2015). Despite setbacks, overall results for the hedgehog inhibitor cyclopamine and its analogs are positive. Due to the complex nature of cancer metastasis, more clinical trials are required

before abandoning this potential treatment. All of these cyclopamine derivatives, however, require cyclopamine as starting material, so the need for cyclopamine supply is still great.

Recently, another structurally unrelated Hh antagonist, GDC-0449 (Vismodegib), was approved by the FDA for use in adults with basal cell carcinoma. This is the first drug approved for cancer treatment by perturbation of the Hh pathway. Mutations in cancer cells are common, and multiple treatment options are required. Other cyclopamine analogs and unrelated Hh inhibitors have been in development (Ma *et al.* 2013). The FDA approval of GDC-0449, the continued synthesis of more stable, active, and biologically available Hh inhibitors, and the ongoing testing of IPI-926 supports cancer therapy by Hh signaling inhibition and highlights the need for more research in this field.

1.4 Cyclopamine supply

The promising activity of cyclopamine and its derivatives for cancer treatment is complicated by the limited supply. The natural source, *V. californicum*, is a slow growing plant for which reliable and robust propagation techniques are yet to be established (Song *et al.* 2014). Improved extraction techniques have been developed, but issues include substantial benzene consumption in addition to the limited supply of material. Total synthesis is difficult and inefficient, with the most promising technique yielding only 1% cyclopamine. If cyclopamine or any of its derivatives are approved for medicinal use, the current supply will not meet the demand (Heretsch *et al.* 2010). Understanding how *V. californicum* produces cyclopamine is one way to propel industrial production. Once the genes producing enzymes that are involved in the chemical transformations of cyclopamine are discovered, they can be introduced into another, more suitable, organism for production.

1.5 Current knowledge of steroid alkaloid biosynthesis

Little is known about the biosynthesis of steroid alkaloids, especially at the gene level. Early studies performed in Japan during the 1970's lead by Kaneko et al. provided a framework for steroid alkaloid biosynthesis (including cyclopamine) in *V. grandiflorum,* but no genes were identified (Figure 7). Possible intermediates for cyclopamine, including 22, 26-dihydroxycholesterol (dormantinol), 22-keto-26-hydroxycholesterol (dormantinone), verazine, and solanidine, were established based upon structure and their presence in the plant (Kaneko *et al.* 1978, Kaneko *et al.* 1979, Kaneko *et al.* 1970a, Kaneko *et al.* 1970b, Kaneko *et al.* 1975, Kaneko *et al.* 1977). Acetate and cholesterol were also established as early precursors by feeding their 14C labeled counterparts to *V. grandiflorum* and monitoring incorporation into

steroid alkaloids (Kaneko *et al.* 1970b). In addition, L-arginine was identified as the source of nitrogen in the piperidine ring. The accumulation of L-arginine in the rhizome followed by a significant decrease over time correlated with the accumulation of verazine(Kaneko *et al.* 1976). Mechanistically, the closing of the E-ring is hypothesized to occur when the lone pair of electrons on the nitrogen of teinemine undergoes a nucleophilic attack on the C-16 in the presence of a suitable leaving group. Another experiment revealed the conversion between the solanidine and jerveratrum alkaloids may require light. Etiolated *V. grandiflorum* plants accumulated radioactive solanidyl glycoside when grown in the dark. When illuminated, they started to accumulate radioactive jerveratrum alkaloids. In addition, this information supports a Wagner-Meerwein rearrangement, which would require ATP (Heretsch *et al.* 2010, Kaneko *et al.* 1979, Kaneko *et al.* 1972). Interestingly, one study based on the biosynthesis of solasodine in *Solanum laciniatum* indicates the addition of nitrogen directly follows hydroxylation at the C-26 position, then possibly functionalization at positon 22 (Tschesche and Brennecke 1980). This contradiction, however, could be attributed to different pathways in the two species.

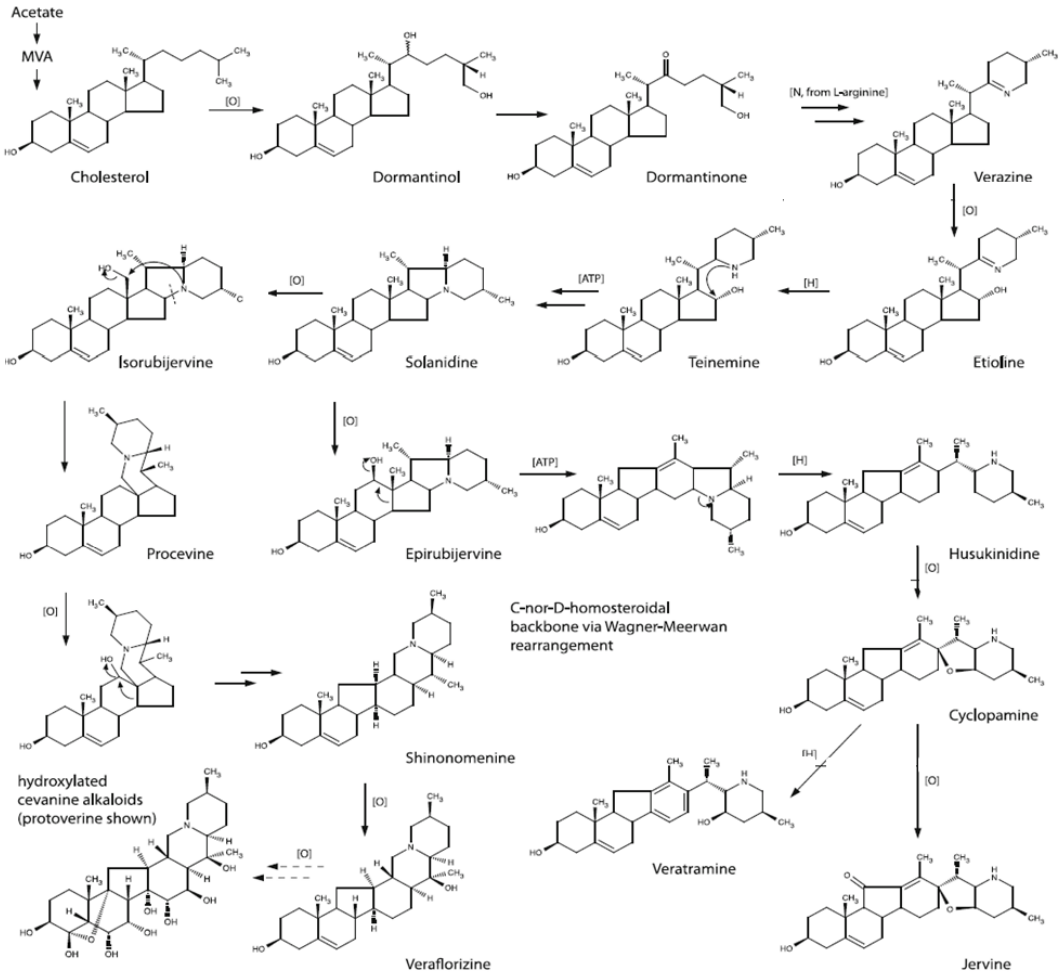

Figure 7. Proposed biosynthetic pathway of steroid alkaloids in *Veratrum grandiflorum*. Adapted from Chandler and McDougal 2014.

More recently, work on the biosynthesis of α-tomatine, a spirosolane steroid alkaloid from *S. lycopersicum,* has provided some insight on steroid alkaloid biosynthesis at the gene level. Researchers used RNAi to downregulate potential biosynthetic genes (based on expression correlation studies) and confirmed their functionality by analyzing accumulation of intermediates. They found that hydroxylation of cholesterol at position C-22 was followed by hydroxylation at position C-26 and that these reactions involved cytochrome P450 (CYP450) enzymes. Several steps followed to form the E-ring which involves, at least, a CYP450 and a 2-oxoglutarate-dependent dioxygenase. Oxidation at position C-26 was also catalyzed by a CYP450 followed by addition of an amino group by a transaminase. Members of Solanaceae do not produce

jerveratrum alkaloids containing the C-*nor*-D-*homo* ring system, however, the study does provide information on enzyme classes involved and potential order of transformations.

1.6 Hypothesis and proposal

The issue remains that all genes involved in the biosynthesis of cyclopamine are unknown. Knowledge of the underlying genes is essential for bioengineering the pathway into another, more suitable, biological production system. Our goal is to discover the genes in the biosynthetic pathway to cyclopamine and engineer the pathway into the emerging oilseed crop *Camelina sativa*, chosen for its robust growth, low input requirements, and ease of transformation (Bansal and Durrett 2015). We hypothesize that correlating a cyclopamine accumulation model (based upon the concentration of cyclopamine in different *V. californicum* tissues) with gene expression (obtained by Illumina high-throughput sequencing); we can reduce the number of candidate genes to an acceptable level for functional characterization. Candidate genes will be selected based upon correlation with the cyclopamine accumulation model in conjunction with predicted functional annotation. Genes will be expressed in an Sf9 insect cell culture system to produce enzymes for testing. Functional enzymes will then be transformed into *C. sativa* and driven by seed specific promoters in order to accumulate *V. californicum* metabolites in seed. This will be the first attempt to discover cyclopamine biosynthetic enzymes and to engineer *C. sativa* for medicinally useful metabolites.

Chapter 2: Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: production of verazine in Sf9 cells (Augustin *et al.* 2015b)

Megan M. Augustin[1], Dan R. Ruzicka[1,13], Ashutosh K. Shukla[1,14], Jörg M. Augustin[1], Courtney M. Starks[2], Mark O'Neil-Johnson[2], Michael R. McKain[1], Bradley S. Evans[1], Matt D. Barrett[3,4], Ann Smithson[3,4], Gane Ka-Shu Wong[5-7], Michael K. Deyholos[8], Patrick P. Edger[9,10], J. Chris Pires[9], James H. Leebens-Mack[11], David A. Mann[12,15], and Toni M. Kutchan[1]

1. Donald Danforth Plant Science Center, St. Louis, Missouri, USA
2. Sequoia Sciences, St. Louis, Missouri, USA
3. Botanic Gardens and Parks Authority Kings Park and Botanic Garden, West Perth, Australia
4. School of Plant Biology, University of Western Australia, Perth, Australia
5. Department of Biological Sciences, University of Alberta, Edmonton AB, Canada
6. Department of Medicine, University of Alberta, Edmonton AB , Canada
7. BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China
8. University of Alberta, Edmonton, Alberta, Canada
9. Bond Life Sciences Center, Division of Biological Sciences, University of Missouri, Columbia, Missouri, USA
10. Department of Plant and Microbial Biology, University of California, Berkeley, California, USA
11. Franklin College of Arts and Sciences, University of Georgia, Athens, Georgia, USA
12. Infinity Pharmaceuticals, Cambridge, Massachusetts, USA

[13]Monsanto Company, 700 Chesterfield Parkway West, St Louis, MO 63017
[14] CSIR-Central Institute of Medicinal and Aromatic Plants, P.O. CIMAP, Lucknow 226015, Uttar Pradesh, India
[15] Cellular Dynamics International, 525 Science Drive, Madison, WI 53711

2.1 Summary

Steroid alkaloids have been shown to elicit a wide range of pharmacological effects that include anticancer and antifungal activities. Understanding the biosynthesis of these molecules is essential to bioengineering for sustainable production. Herein, we investigate the biosynthetic pathway to cyclopamine, a steroid alkaloid that shows promising antineoplastic activities. Supply of cyclopamine is limited, as the current source is solely derived from wild collection of the plant *Veratrum californicum*. To elucidate the early stages of the pathway to cyclopamine, we interrogated a *V. californicum* RNA-seq dataset using the cyclopamine accumulation profile as a predefined model for gene expression with the pattern-matching algorithm Haystack. Refactoring candidate genes in Sf9 insect cells led to discovery of four enzymes that catalyze the first six steps in steroid alkaloid biosynthesis to produce verazine, a predicted precursor to cyclopamine. Three of the enzymes are cytochromes P450 while the fourth is a γ-aminobutyrate transaminase; together they produce verazine from cholesterol.

2.2 Significance statement

The first four enzymes involved in the biosynthesis of the steroid alkaloid verazine, a predicted precursor to the antineoplastic cyclopamine in *Veratrum californicum,* were discovered: cholesterol 22-hydroxylase, 22-hydroxycholesterol 26-hydroxylase/oxidase, 22-hydroxycholesterol-26-al transaminase, and 22-hydroxy-26-aminocholesterol 22-oxidase. The pathway to verazine has been refactored in *Spodoptera frugiperda* Sf9 cells.

2.3 Author contributions

D.A.M. collected *V. californicum* plant material. M.M.A isolated *V. californicum* RNA. D.R.R. analyzed transcriptome data and produced the candidate gene list. D.R.R., A.K.S, and M.M.A. cloned candidates into baculovirus expression vectors and expressed them in Sf9 cells. J.M.A. and M.M.A. developed and performed enzyme assays to determine enzymatic activity and purified metabolites for structure elucidation. C.M.S. and M.O.J. purified metabolites and determined structures by NMR. B.S.E analyzed metabolites by LTQ-Velos Pro Orbitrap. M.R.M. performed deep phylogenetic analysis of cytochromes P450. J.C.P., P.P.E., M.D.B., A.S., G.K.W., and M.K.D. contributed plant materials and RNAseq data from the 1KP project. M.M.A. and T.M.K. analyzed data and wrote the manuscript, which was reviewed by all authors. T.M.K. directed the study.

More specifically, I (M.M.A.) cloned contigs 13284 and 12709 and all of their respective homologs, and homologs of 2646 (originally cloned by D.R.R.). I also cloned additional genes

as indicated in Table S3 and verified the sequences of all genes by DNA sequencing. I performed homologous recombination of genes 13284 (and the respective homologs), two homologs of 12084, and one homolog of 2646 with baculovirus viral DNA in addition to those indicated in Table S3. I performed all Sf9 cell maintenance and infection for enzyme production. I conducted all enzyme assays and Sf9 extracts, ran all instruments and analyzed all data for enzyme function and product production. I produced the data for and assembled the images for figures 1,2,4,5 and supplemental figures S1-S8, S10-S15, and S17. I produced or assembled information for tables S5-S7, and S11-S13. I wrote and edited this manuscript with the help of Toni M. Kutchan.

2.4 Introduction

Plants synthesize numerous specialized metabolites that are used in medicines today (Mishra and Tiwari 2011). Their endogenous function is thought to play a role in communication of the plant with its environment, as these compounds possess an array of biological activities (Mithofer and Boland 2012, Mizutani 2012). An understanding of how these molecules are formed serves a dual role: enabling study of the *in planta* function, as well as development of synthetic biology production platforms for source plants that are not amenable to cultivation or genetic manipulation.

In seeking medicines from natural sources, plants synthesizing the nitrogen containing alkaloids have proven to be a vital source. Of particular pharmaceutical interest are the steroid alkaloids. Members of the families Liliaceae, Apocynaceae, Buxaceae, and Solanaceae are a rich source of these alkaloids. The structural similarity of steroid alkaloids to mammalian steroid hormones may be responsible for their therapeutic effects (Cordell 1998, Jiang *et al.* 2005, Zhou 2003). Our study focuses on the steroid alkaloid cyclopamine from the Liliaceae family. Members of Liliaceae contain three sub-classes of steroidal alkaloids: the jerveratrum-type, the cerveratrum-type, and the solanidine-type (Cordell 1998). Cyclopamine, also known as 11-deoxojervine, is a jerveratrum-type alkaloid that exhibits potent pharmacological properties.

Cyclopamine was originally discovered due to its teratogenic effect that resulted in craniofacial anomalies (cyclopia) in lambs born at high elevation in the northwestern United States in the middle of the 20th century (Keeler 1970a, Keeler 1970b, Keeler and Binns 1966b). Studies supported by the United States Department of Agriculture found that the California corn lily *Veratrum californicum* was the source of the teratogen in the natural diet of the pregnant ewes (Keeler 1968, Keeler 1969, Keeler 1970a). Cyclopamine was discovered to inhibit the

Hedgehog signaling pathway by direct binding to the G protein-coupled receptor Smoothened (Chen *et al.* 2002). As such, cyclopamine has shown promising antineoplastic activities against several cancers in which Hedgehog signaling malfunction is implicated, including pancreatic cancer, renal cell carcinoma, medulloblastoma, basal cell carcinoma, and leukemia (Bahra *et al.* 2012, Behnsawy *et al.* 2013, Berman *et al.* 2002, Gailani *et al.* 1996, Lin *et al.* 2010, Olive *et al.* 2009, Taipale *et al.* 2000). A semi-synthetic analog of cyclopamine, IPI-926, has been in clinical trials for treatment of several cancers including pancreatic cancer and leukemia (Lin *et al.* 2010, Olive *et al.* 2009, Tremblay *et al.* 2009). Due to complicated total synthesis, wild-collected *V. californicum* is the current commercial source of cyclopamine. Commercial cultivation of the plant has not yet been successful. Cyclopamine is thereby an attractive target for biotechnological production. The first obstacle to this approach is, however, obtaining knowledge of the underlying biosynthetic genes.

Cyclopamine biosynthesis is believed to begin with cholesterol, a common precursor to most steroidal alkaloids (Cordell 1998, Kaneko *et al.* 1970b). Early studies on species of the Liliaceae and Solanaceae families have proposed transformation of cholesterol to metabolites identified in plant extracts that include dormantinol, dormantinone, verazine, and etioline. All of these molecules are proposed intermediates in the biosynthesis of cyclopamine (Chandler and McDougal 2014, Kaneko *et al.* 1977). Previous feeding studies shed limited light on pathway order and suggested that L-arginine is the origin of the nitrogen atom (Kaneko *et al.* 1977, Kaneko *et al.* 1976, Tschesche and Brennecke 1980). A more recent report on the biosynthesis of Solanaceous alkaloids utilizing gene co-expression analysis and RNAi knock-out transgenic plants in *Solanum tuberosum* and *Solanum lycopersicum* outlines a biosynthetic pathway leading to the synthesis of the spirosolane-type steroid alkaloid α-tomatine. The enzyme classes identified in the *Solanum* pathway include several cytochrome P450 enzymes, a transaminase, and a 2-oxoglutarate-dependent dioxygenase (Itkin *et al.* 2013). Questions remain, however, concerning the formation of steroid alkaloids in *Veratrum* as no genes or enzymes have been reported.

Biochemical pathway elucidation in non-model systems has historically taken decades to complete, but bioinformatic technologies are revolutionizing the approach. A prominent example of former methods used in pathway discovery is morphine biosynthesis in opium poppy, *Papaver somniferum*. Though morphine was discovered in the early 1800's, the biosynthetic pathway is still incomplete at the gene level. Genes encoding only 6 of the 8 enzymes committed to the biosynthesis of morphine have been isolated and characterized from the

1990's to the present – several decades of work to uncover fewer than 8 genes (Gesell *et al.* 2009, Grothe *et al.* 2001, Hagel and Facchini 2010, Unterlinner *et al.* 1999, Ziegler *et al.* 2006). Now, high-throughput sequencing technology enables new approaches to biochemical pathway discovery in the non-model system by providing nucleotide sequence data acquisition at a previously unparalleled rate. A combination of bioinformatics and high-throughput sequencing has the potential to shorten natural product pathway discovery if the challenges of interrogating such large datasets from non-model systems is overcome. We present herein the discovery of four *V. californicum* enzymes that catalyze the first six steps from cholesterol to verazine, a predicted precursor to the steroid alkaloid cyclopamine. We utilize a biosynthetic gene discovery method founded on correlating metabolite accumulation with RNA-seq generated gene expression data. By stepwise refactoring of the pathway in *Spodoptera frugiperda* Sf9 cells, we eliminate the need to synthesize biosynthetic intermediates for validation of pathway enzyme activity.

2.5 Results

2.5.1 RNA-seq and de novo transcriptome assembly

Multiplex paired-end sequencing of *V. californicum* cDNA produced from bulb, flower, leaf, fall rhizome, spring rhizome, fall root, green shoot, white shoot, and tissue culture samples on two 2 x 50 bp Hi-Seq channels produced over 201 million raw reads which were analyzed and filtered for artifacts/contaminants. The *de novo* short read assembly generated 56,994 contigs with a median length of 323 bp and an N50 of 1,471 bp (refer to Table S1 and Table S2 for Illumina reads generated per file and transcriptome statistics, respectively). Information on accessing the raw reads and assembled transcriptome can be found in Data S1. Raw mapped read-counts were utilized as a metric of relative gene expression. The average contig sequence length indicates high quality assembly and was sufficient for downstream sequence alignment and phylogenetic gene tree estimation.

2.5.2 Transcriptome dataset interrogation

To provide annotation, predicted peptide sequences were queried against the Pfam and Superfamily databases in addition to BLAST searches at NCBI using the non-redundant protein sequence database. Expression data for each contig was normalized using total reads per organ type to serve as the dataset for Haystack (http://haystack.mocklerlab.org/)(Mockler *et al.* 2007).

LC-MS/MS determination of the steroid alkaloid profile in the same *V. californicum* tissues used for RNA-seq identified a pronounced accumulation of cyclopamine in rhizome, followed by root and bulb (Figure S1), suggesting that biosynthesis occurs in underground organs of the plant. Since secondary metabolites are most often synthesized at or near their site of accumulation (Huang and Kutchan 2000, Nims *et al.* 2006, Onrubia *et al.* 2011, Weid *et al.* 2004)**,** we initially hypothesized that underground tissues (rhizome, root, and bulb) in *Veratrum* are biosynthetic for cyclopamine.

Since *ca.* 20 times more cyclopamine accumulates in subterranean organs, these tissues were given a value of 20 for the Haystack input while the above ground organs were designated a value of 1 in order to create a generalized model for cyclopamine biosynthesis. In our approach, the LC-MS/MS alkaloid data for *Veratrum* is the input *model* used to search the deep transcriptome experimental *dataset* of *Veratrum*, correlating metabolite accumulation and gene expression data as previously reviewed (Saito *et al.* 2008). We obtained 3,219 contigs that fit the 20:1 subterranean organ:aerial organ cyclopamine accumulation model.

In addition to identification of genes exhibiting correlating expression patterns with cyclopamine accumulation, the protein-coding gene sequences in the *Veratrum* RNA-seq transcriptome dataset were classified into putative gene families using PlantTribes 2.0 (http://fgp.bio.psu.edu/tribedb/10_genomes/index.pl) (Wall *et al.* 2008). This dataset was used to better define and cluster the *Veratrum* transcripts that are of most interest to the cyclopamine pathway. Cytochromes P450 were chosen from the resulting dataset due to the hypothesized oxidative transformations necessary to convert cholesterol into cyclopamine. The alternative enzyme class would be 2-oxoglutarate-dependent dioxygenases, if the cytochrome P450 dataset did not yield positive results.

Top candidate genes from the Haystack and tribe clustering analysis with the highest homology to cytochrome P450 enzymes involved in triterpene and fatty acid hydroxylation/ biosynthesis were selected for downstream functional characterization (Table S3). Incorporation of nitrogen into the steroid skeleton is required; therefore, aminotransferases fitting the Haystack model were also included in the candidate gene list (Table S4). Full-length candidate cDNAs were expressed in *S. frugiperda* Sf9 cells using a baculovirus-based expression system allowing for the accommodation of multi-virus infections. *S. frugiperda* Sf9 cells provide a facile synthetic biology platform for the systematic refactoring of plant biosynthetic pathways (Diaz Chavez *et al.* 2011)**.**

2.5.3 Cholesterol 22-hydroxylase

The top-scoring candidate cDNAs resulting from interrogation of the *V. californicum* transcriptome dataset were systematically and individually introduced, together with *Eschscholzia californica* cytochrome P450 reductase (CPR), into *S. frugiperda* Sf9 insect cells, which were harvested as previously described and used in enzyme assays with cholesterol as substrate (Diaz Chavez *et al.* 2011, Rosco *et al.* 1997). Most cytochrome P450 enzymes require the addition of a reductase for heterologous expression. In our hands, differences in enzyme activity were not dependent on the plant species from which the CPR was derived (Gesell *et al.* 2009). Cholesterol was chosen as the initial precursor for study based upon existing knowledge of steroid alkaloid biosynthesis (Cordell 1998, Kaneko *et al.* 1970b). Related compounds were also tested to determine enzyme specificity (Table S5). The contig designated 2646, which is annotated as a steroid C-22 hydroxylase, added a hydroxyl group to the 22-position of cholesterol exclusively in the *R* orientation as determined by GC-MS (Figure 1; Figure S2 a; Figure S3 a). The hydroxylation demonstrated in Figure 1 used extracts of Sf9 cells co-expressing contig 2646 and *E. californica* CPR as we have discovered their ability to utilize endogenous Sf9 cholesterol as substrate. One variant of 2646 was identified by RT-PCR having 99.8% identity and catalyzing the same enzymatic reaction. Cytochrome P450 (CYP) assignments for both homologs are CYP90B27v1 and CYP90B27v2, where version CYP90B27v1, accession KJ869252, was chosen for subsequent analysis. We have designated this enzyme cholesterol 22-hydroxylase. CYP90B27 also hydroxylated 26-hydroxycholesterol (27(25*R*)-hydroxycholesterol) and 7β-hydroxycholesterol in the 22-position as determined by enzyme assay, but was unable to hydroxylate the 24-methylsterol campesterol, a metabolite leading into brassinosteroid biosynthesis (Figure S2 e; Figure S4; Figure S5). CYP90B27 oxidizes the hydroxyl group at the 22-position to a ketone, but only to a low degree (Figure 1). The identity of the enzymatic product of CYP90B27 acting on cholesterol was confirmed by comparison to 22(*R)*-hydroxycholesterol standard. The diastereomers 22(*R*)- and 22(*S*)-hydroxycholesterol are chromatographically resolved by the GC-MS method used (Figure S6).

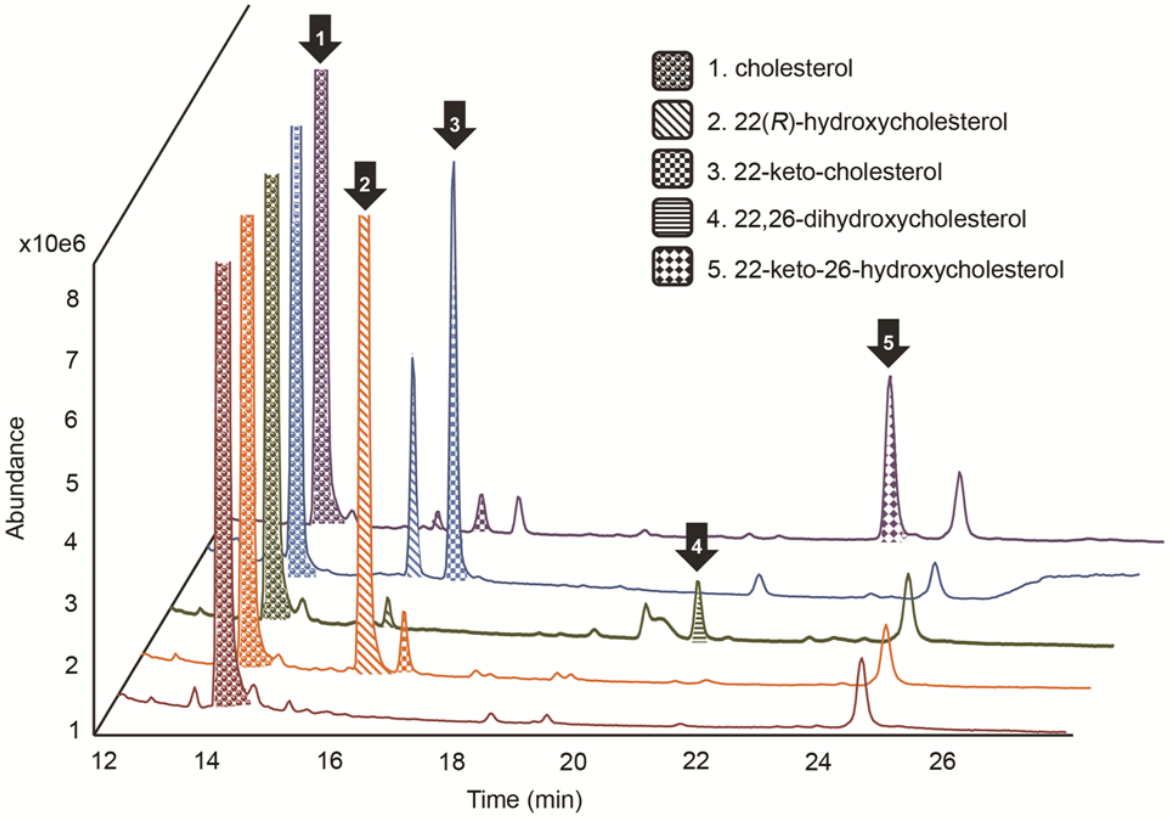

Figure 1. GC-MS Overlay of *S. frugiperda* Sf9 extracts expressing *Veratrum californicum* genes. Extracts of *S. frugiperda* Sf9 cells infected with varying combinations of baculovirus containing genes from *V. californicum* were extracted and analyzed by gas chromatography mass spectrometry. Each colored chromatograph corresponds to the following: **Red**-CYP719A14 (control cytochrome P450) and CPR, **Orange**-CYP90B27 (cholesterol 22-hydroxylase) and CPR, **Green**-CYP90B27, CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), and CPR, **Blue**-CYP90B27, CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), and CPR, **Purple**-CYP90B27, CYP94N1, CYP90G1, and CPR. Metabolites are numbered according to the legend and shaded for clarity. CPR refers to the cytochrome P450 reductase from *Eschscholzia californica* and control P450 refers to CYP719A14 cheilanthifoline synthase from *Argemone mexicana*.

2.5.4 22-Hydroxycholesterol 26-hydroxylase/oxidase

To identify the second enzyme in the pathway, a series of triple infections of *S. frugiperda* Sf9 cells were carried out that contained CYP90B27 and *E. californica* CPR, but varied the second enzyme. Candidates for the second enzyme were the remaining top-scoring candidate cDNAs resulting from interrogation of the *V. californicum* transcriptome dataset. Contig 12709, which annotated as a fatty acid hydroxylase, was found to hydroxylate 22(*R*)-hydroxycholesterol at the C-26 position forming 22,26-dihydroxycholesterol by GC-MS (Figure 1; Figure S4). This enzyme also oxidizes the hydroxyl group at the 26 position creating a highly reactive 22-

hydroxycholesterol-26-al as detected by LC-MS/MS (Figure S7 a). Four variants were generated by RT-PCR with identities ranging from 93–99%; all catalyze the same reaction. Hydroxylation of cholesterol by 12709 was not detected (Figure S3); this enzyme was therefore designated 22-hydroxycholesterol 26-hydroxylase/oxidase. CYP assignments are CYP94N1v1, CYP94N1v2, CYP94N2v1, and CYP94N2v2. In figures and subsequent text, CYP94N1 refers to accession KJ869255 which was chosen as the representative variant for this study. The identity of 22,26-dihydroxycholesterol produced by action of 12709 on 22(*R*)-hydroxycholesterol was ultimately determined using the 22-hydroxylating activity of CYP90B27 to produce 22,26–dihydroxycholesterol from pure 26-hydroxycholesterol and comparing the mass spectra of the two enzymatic products (Figure S4).

2.5.5 22-Hydroxycholesterol-26-al transaminase

To identify the third enzyme in the pathway, a series of quadruple infections of *S. frugiperda* Sf9 cells were carried out that contained CYP90B27, CYP94N1, and *E. californica* CPR, but varied the third enzyme. Candidates for the third enzyme were the remaining top-scoring candidate cDNAs. A γ-aminobutyric acid (GABA) transaminase (contig 12084) was shown to incorporate nitrogen into the 26-position of 22-hydroxycholesterol-26-al using GABA as an amino group donor to produce 22-hydroxy-26-aminocholesterol as detected by LC-MS/MS (Figure S8 a). Three variants were found by RT-PCR, each with over 99% identity and all able to catalyze the same reaction. Evidence for the structure of 22-hydroxy-26-aminocholesterol was obtained by high resolution MS (Figure S9). 22-Hydroxy-26-aminocholesterol was not produced in Sf9 enzyme assays that included CYP90B27, CYP90G1 (see below), GABAT1, and CPR but lacked CYP94N1 (Figure S7 b; orange), indicating that addition of nitrogen to position 26 is likely. This enzyme was designated 22-hydroxycholesterol-26-al transaminase. In figures and subsequent text, GABAT1 refers exclusively to accession KJ89263, the variant chosen for subsequent study.GABAT1 was expressed and purified from *E. coli* and used in enzyme assays to test potential co-substrates GABA, L-arginine, and L-glutamine (Figure S10). GABA was the only accepted amino group donor. Minor peaks in the L-arginine and L-glutamine assays are likely due to GABA present in *E. coli* (Dhakal *et al.* 2012).

2.5.6 22-Hydroxy-26-aminocholesterol 22-oxidase

To identify the fourth enzyme in the pathway, a series of quintuple infections of *S. frugiperda* Sf9 cells were carried out that contained CYP90B27, CYP94N1, GABAT1, and *E. californica* CPR, but varied the fourth enzyme. Candidates for the fourth enzyme were the remaining top-scoring candidate cDNAs. Contig 13284 was also annotated as a steroid C-22 hydroxylase, but hydroxylation at the 22-position of cholesterol was not readily detected. It was, however, able to oxidize an existing hydroxyl group at position 22 to a much greater degree than CYP90B27 as shown by GC-MS (Figure 1; Figure S2 b). 13284 oxidizes the 22-hydroxy position of 22(*R*)-hydroxycholesterol to form 22-keto-cholesterol (Figure 1), 22,26-dihydroxycholesterol to form 22-keto-26-hydroxycholesterol (Figure 1; Figure S8 b), and 22-hydroxy-26-aminocholesterol to form a short lived intermediate that cyclizes to verazine (Figure 2; Figure S8 c). Four variants were isolated using RT-PCR, each having more than 97% identity and all able to perform the same reaction. CYP designations for each sequence are: CYP90G1v1, CYP90G1v2, CYP90G1v3, and CYP90G2. In figures and subsequent text, CYP90G1 refers to accession KJ869260 which was selected for further analysis. The structures of the enzymatic products 22-keto-cholesterol and 22-keto-26-hydroxycholesterol were confirmed by NMR spectroscopy (Data S2). The structure of verazine was supported by high resolution MS (Figure S9). The enzyme was subsequently designated 22-hydroxy-26-aminocholesterol 22-oxidase. Accession numbers and designations for all characterized enzymes can be found in Table S6.

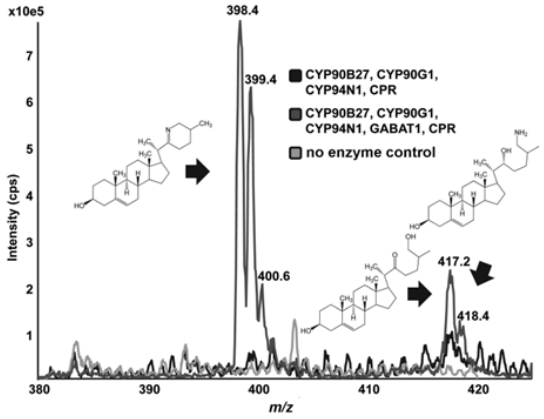

Figure 2. Production of verazine by heterologous expression of *V. californicum* genes in *S. frugiperda* Sf9 cells. Select genes were introduced into *S. frugiperda* Sf9 cells using a baculovirus expression system. Metabolites were extracted and analyzed by LC-MS/MS in the full scan Enhanced MS mode detecting 380 – 425 *m/z*. Each chromatogram represents the combination of genes as follows: **Black**-CYP90B27 (cholesterol 22-hydroxylase), CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), and CPR; **Dark Grey**-CYP90B27, CYP94N1, CYP90G1, GABAT1 (22-hydroxycholesterol-26-al transaminase), CPR; **Light Grey**-no enzyme control. CPR refers to the cytochrome P450 reductase from *E. californica*, and GABAT1 refers to the γ-aminobutyric acid transaminase 1 from *V. californicum*. Peak at 398.4 is verazine, peak at 417.2 is 22-keto-26-hydroxycholesterol, and peak at 418.4 is for 22-hydroxy-26-aminocholesterol.

2.5.7 Biosynthetic pathway to verazine

The substrate specificities determined for these enzymes suggested a potential metabolic grid in the metabolism of

cholesterol. CYP90B27 catalyzes the 22-hydroxylation of cholesterol; this is most likely the first step in the biosynthesis of steroid alkaloids in *V. californicum*, confirmed by the inability of CYP94N1 and CYP90G1 to accept cholesterol as substrate (Figure S3).

To establish the pathway order after 22-hydroxylation of cholesterol, a series of enzyme assays were carried out using *S. frugiperda* Sf9 cell extracts containing each individual candidate enzyme. The order of addition for each enzyme was varied, and products were analyzed by GC-MS or LC-MS/MS. The flow chart for both sets of experiments is presented in Figure S11. As seen in Figure S12 c and e, 22-keto-26-hydroxycholesterol was only produced at detectable levels by CYP90G1 from 22,26-dihydroxycholesterol. CYP94N1 was unable to hydroxylate 22-keto-cholesterol at levels detected by GC-MS. The ability of CYP90G1 to accept 22,26-dihydroxycholesterol as substrate and produce 22-keto-26-hydroxycholesterol, along with the lack of product detection for CYP94N1 incubated with 22-keto-cholesterol, provided evidence that CYP94N1 acted directly after CYP90B27. This evidence was substantiated with another set of enzyme assays, beginning with CYP90B27, showing that CYP94N1 produced little product when provided with 22-keto-cholesterol (using the increased sensitivity of LC-MS/MS for detection) as seen in Figure S7 b; blue, as compared to the large amount of product produced when 22,26-dihydroxycholesterol is acted upon by CYP90G1 (Figure S7 a; blue).

22-Hydroxy-26-aminocholesterol was produced in Sf9 cells by CYP90B27, CYP94N1, GABAT1, and CPR in the presence or absence of CYP90G1 (Figure S8 a; c). GABAT1, therefore, did not require a 22-ketone moiety on the substrate. When CYP90B27 acted in the presence of CYP94N1, several side products were detected in addition to 22,26-dihydroxycholesterol (Figure S7 a; red). These compounds included 22-keto-26-hydroxycholesterol, 22-hydroxycholesterol-26-al and two unidentified products. Since an amino group was not added to the 22-ketone moiety of 22-keto-26-hydroxycholesterol, 22-keto-26-hydroxycholesterol most likely does not participate in this steroid alkaloid pathway. The short lived and highly reactive 22-hydroxycholesterol-26-al must therefore be the substrate of GABAT1. After an amino group is transferred to the C-26 aldehyde, CYP90G1 oxidizes the C-22-hydroxyl moiety to a ketone, and cyclization to verazine occurs (Figure 2; Figure S8 c).

Evidence for the short-lived intermediate 22-hydroxycholesterol-26-al was obtained by dimedone aldehyde trapping (Figure S13). The amino group is added prior to oxidation of the C-22 hydroxyl group; therefore, the amino group must be transferred to a C-26 aldehyde. The structure of the predicted cyclic imine verazine was supported by borohydride reduction of the

double bond (Figure S14) and exact mass analysis as demonstrated by high resolution MS (Figure S9).

The biosynthetic pathway proposed herein (Figure 3) is similar to a hypothesized pathway presented in earlier studies of steroid alkaloids in the genus *Veratrum.* We found, however, that 22-keto-26-hydroxycholesterol (dormantinone) was not readily accepted as substrate for GABAT1, as verazine was not formed when supplied this compound (Figure S7 a; purple). In contrast, verazine was formed when 22-hydroxy-26-amino-cholesterol was supplied to CYP90G1 (Figure S7 a; green). In further support of our proposed pathway, the biosynthetic intermediates 22-hydroxy-26-aminocholesterol, 22-keto-26-hydroxycholesterol, and verazine were detected in *V. californicum* extracts by LC-MS/MS (Figure S15) and the distribution of these intermediates followed the same accumulation pattern as cyclopamine (Figure S1). 22,26-Dihydroxycholesterol (dormantinol) and verazine have also been previously detected in steroid alkaloid producing *Veratrum* species and hypothesized to be intermediates in steroid alkaloid biosynthesis (Adam *et al.* 1967, Kaneko *et al.* 1977)**.**

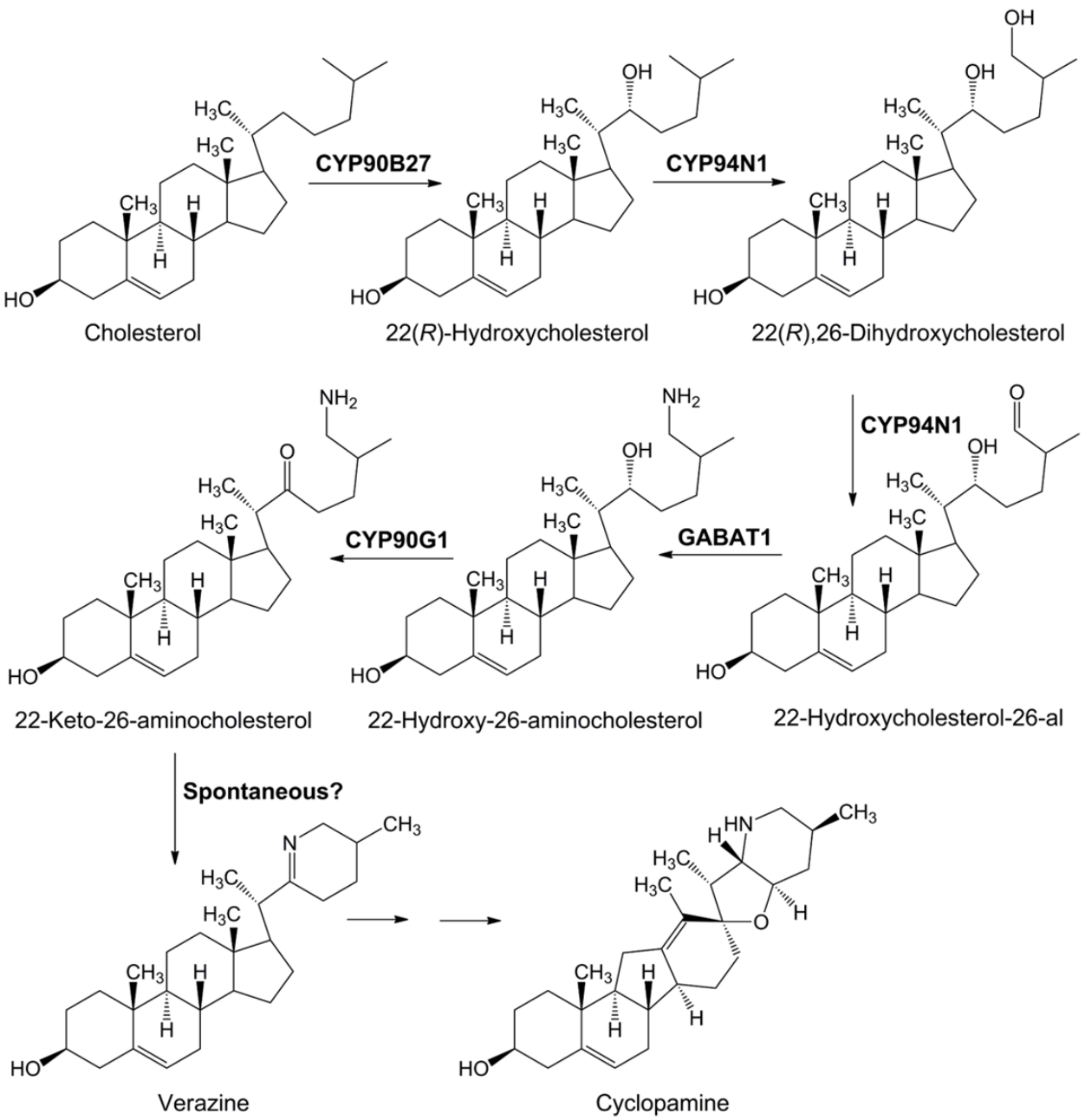

Figure 3. Proposed *V. californicum* cyclopamine biosynthetic pathway leading from cholesterol. Cholesterol is first hydroxylated at position C-22 in the *R*-orientation by CYP90B27 (cholesterol 22-hydroxylase), followed by hydroxylation/oxidation at position C-26 by CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase). Next, a transamination reaction by GABAT1 (22-hydroxycholesterol-26-al transaminase) transfers an amino group from γ-aminobutyric acid to the C-26-aldehyde, forming 22-hydroxy-26-aminocholesterol. The C-22-hydroxy group is then oxidized to a ketone by CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase) to form 22-keto-26-aminocholesterol, a reactive intermediate that cyclizes to verazine. GABAT1 refers to the γ-aminobutyric acid transaminase 1 from *V. californicum*.

2.5.8 Site of steroid alkaloid biosynthesis in V. californicum

A comparison was made between biosynthetic gene expression profiles and cyclopamine accumulation in *V. californicum* (Figure 4). A pattern emerged that indicates biosynthetic genes are most highly expressed in root, bulb, and spring rhizome, while cyclopamine (Figure 4; Figure S1) is highest in spring- and fall rhizome. The hypothesized intermediates 22-hydroxy-26-aminocholesterol, 22-keto-26-hydroxycholesterol, and verazine follow a similar trend with highest accumulation in root, spring- and fall rhizome (Figure S15). Moreover, the higher level of cyclopamine in fall rhizome compared to spring rhizome suggests an accumulation of the steroid alkaloid during rhizome growth in summer (Figure S1). Interestingly, fall rhizome has significantly more cyclopamine relative to transcript level; the opposite is true for bulb (Figure 4).

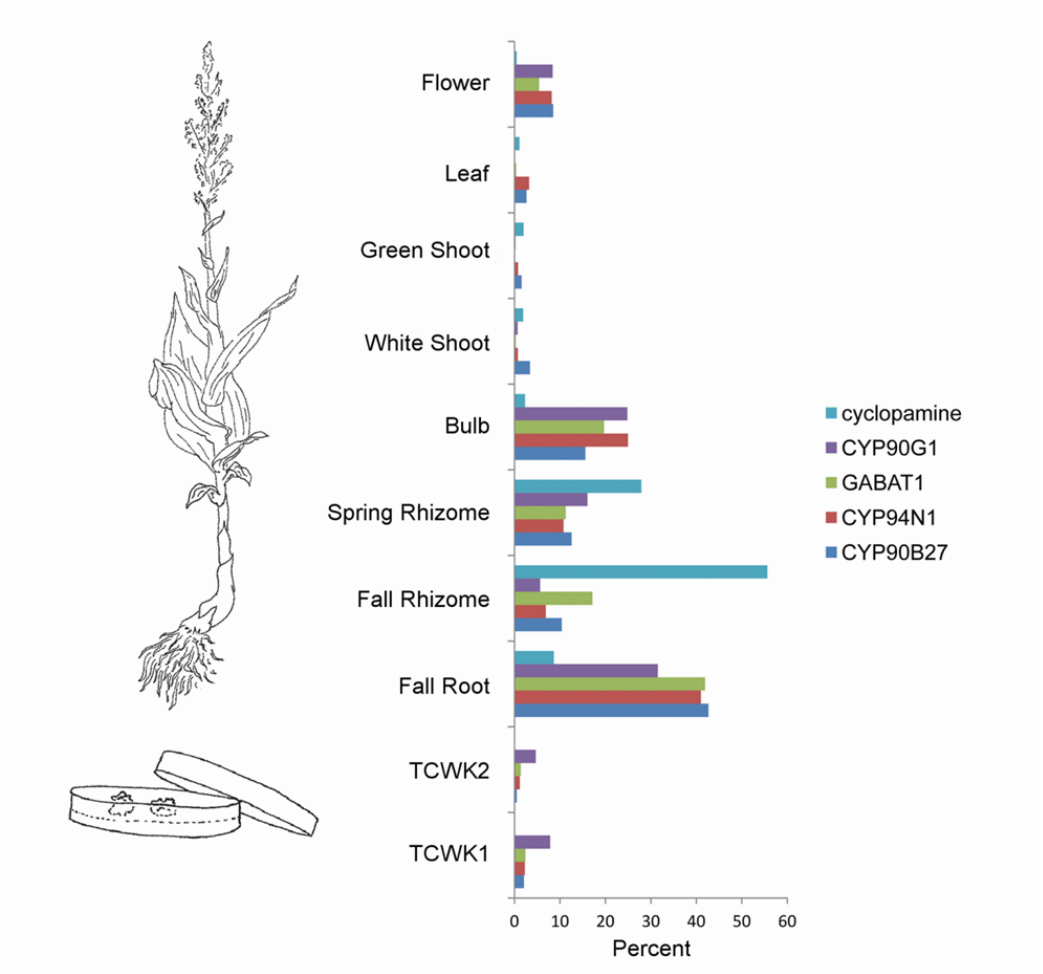

Figure 4. Comparison of cyclopamine accumulation and gene expression of steroid alkaloid biosynthetic genes. Tissues from *V. californicum* were extracted and analyzed by LC-MS/MS for cyclopamine quantitation. Transcript abundance was analyzed by alignment of individual reads to the assembled transcriptome for gene expression. Both gene expression and cyclopamine accumulation are shown as a percent of the total for comparison. The abbreviations TCWK1 and TCWK2 stand for tissue culture one- and two weeks after transfer to fresh media (respectively). CYP90G1 refers to 22-hydroxy-26-aminocholesterol 22-oxidase, GABAT1 refers to 22-hydroxycholesterol-26-al transaminase (γ-aminobutyric acid transaminase 1 from *V. californicum*), CYP94N1 refers to 22-hydroxycholesterol 26-hydroxylase/oxidase, and CYP90B27 refers to cholesterol 22-hydroxylase.

2.6 Discussion

The evolutionary origin of specialized metabolism like that in the *Veratrum* genus is of particular interest not only for understanding species divergence but also for the recognition of additional species containing related compounds with potential therapeutic value. Biochemical analysis of the *Veratrum* steroid alkaloid pathway herein alongside a recently proposed biosynthetic pathway for steroid glycoalkaloids in *Solanum lycopersicum* (Itkin *et al.* 2013) suggests

independent origins of the *Solanum* steroid alkaloids and the *Veratrum* steroid alkaloids. In *S. lycopersicum*, initial transformations of cholesterol include C-22 hydroxylation followed by C-26 hydroxylation and closure of the E-ring. Oxidation at C-26 and subsequent transamination at that position occur without the formation of the intermediate verazine. In conjunction with our results, previous work on steroid alkaloid formation in *Veratrum* does not support E-ring closure prior to aldehyde formation and transamination (Kaneko *et al.* 1970a, Kaneko *et al.* 1975, Kaneko *et al.* 1977, Kaneko *et al.* 1976). Verazine production requires formation of the F-ring following transamination prior to E-ring closure. If the pathway was identical to that proposed in *S. lycopersicum*, E-ring closure prior to amination would not allow for the formation of verazine. Oxidation at C-26 by the multifunctional CYP94N1 may be the gateway reaction when comparing the point of divergence between the two pathways. This oxidation in *Veratrum* allows for the immediate addition of nitrogen, leading to F-ring closure and verazine formation. In contrast, the action of GAME11, the 2-oxoglutarate-dependent dioxygenase in the *Solanum* pathway, in conjunction with other unidentified enzymes after C-26 hydroxylation, leads to closure of the E-ring prior to C-26 oxidation and nitrogen addition. The possibility exists that multiple steroid alkaloid biosynthetic pathways occur in one or both families and those respective genes are yet to be discovered, as solanidine-type and verazine-like compounds have been identified in both members of the Liliaceae and Solanaceae (Cordell 1998).

The independent origin of the two pathways is further corroborated by the phylogenetic relationship of the enzymes involved. A phylogenetic analysis of select cytochrome P450 enzymes, including several involved in steroid metabolism (Figure 5, Table S7), demonstrates this distinction. CYP90B27 in *V. californicum* and CYP72A186 (GAME7), the proposed cholesterol 22-hydroxylase in *S. lycopersicum*, are in distinct pan-angiosperm clades suggesting recruitment from separate P450 subfamilies; likewise for CYP72A208 (26-hydroxylase in *S. lycopersicum*, GAME8) and *V. californicum* CYP94N1. Similarly, CYP88B1 (GAME4), the *S. lycopersicum* enzyme that performs oxidation at position 26, falls into a well-supported clade distinct from the clade containing CYP94N1. The phylogenetic placement of these enzymes alludes to potential convergent evolution of these genes. The similarity of *V. californicum* enzymes CYP90B27 and CYP90G1 with the CYP90B1s, enzymes known to participate in primary brassinosteroid metabolism, may be indicative of steroid alkaloid biosynthesis evolution, in part, deriving from the brassinosteroid pathway.

Figure 5. Phylogenetic tree of select plant cytochrome P450 enzymes. Nucleotide sequences obtained from Genbank, Uniprot, and the Sol Genomics Network of select cytochrome P450 enzymes were aligned by codon with the Muscle algorithm. Cytochrome P450 designations, species, and their corresponding function can be found in Table S13. Only experimentally determined functions are designated in the figure. *V. californicum* enzymes involved in cyclopamine biosynthesis are represented by two stars; *S. lycopersicum* enzymes involved in steroid alkaloid metabolism are represented with one star. Phylogenetic reconstruction was performed using the Maximum Likelihood statistical method with bootstrapping in MEGA version 6.06 with default parameters.

An in-depth phylogenetic analysis (Figure S16, Table S8) of the steroid modifying genes from *V. californicum* and *S. lycopersicum* alongside homologous cytochrome P450 genes taken from transcriptome assemblies generated by the 1KP and MonAToL sequencing projects (Data S1) substantiate the hypothesis of independently derived pathways. Three major cytochrome P450 clades emerge. Both *V. californicum* CYP90B27 and CYP90G1 are localized in Clade 1. Clade 1 contains a number of ancient gene duplications that show evidence of paralog retention with expression of multiple copies within species, especially in monocots. In agreement with Figure 5, SlCYP90B1 falls into the same well-supported subclade as these two *Veratrum* genes with CYP90B27 being more closely related to SlCYP90B1. Clade 2 contains the *S. lycopersicum* genes GAME6, GAME7, and GAME8, but no sampled *Veratrum* genes. Interestingly, this clade shows several eudicot gene duplications, most likely genome-specific tandem duplications that are not observed in the monocot species included in this study. *V. californicum* CYP94N1 is found in Clade 3. Clade 3 appears to be the least diverse of the three P450 clades containing multiple copies of homologs only in a few taxa, potentially related to whole genome duplication events.

We functionally investigated the relatedness of the *V. californicum* GABAT1 that incorporates nitrogen into 22-hydroxycholesterol-26-al and the *S. lycopersicum* GABA transaminase isozyme 2 involved in steroid alkaloid biosynthesis. These enzymes share 64% identity at the amino acid level. Due to the similarities of the two pathways, we tested whether *S. lycopersicum* GABA transaminase isozyme 2 can incorporate nitrogen into 22-hydroxycholesterol-26-al. We demonstrated that *S. lycopersicum* GABA transaminase isozyme 2 was able to convert 22-hydroxycholesterol-26-al to 22-hydroxy-26-aminocholesterol and, in conjunction with CYP90G1, resulting in subsequent cyclization to verazine (Figure S8 d). *S. lycopersicum* GABA transaminase isozyme 2 was used as query to BLAST the *V. californicum* transcriptome; the best hit was another transaminase, contig 674. 674, designated GABAT2, has 68% identity to *S. lycopersicum* GABA transaminase isozyme 2 and 69% identity to *V. californicum* GABAT1 at the amino acid level. Despite the sequence homology to both *V. californicum* GABAT1 and *S. lycopersicum* GABA transaminase isozyme 2, GABAT2 was unable to convert 22-hydroxycholesterol-26-al to 22-hydroxy-26-aminocholesterol at detectable levels (Figure S17).

Overall, our data indicates that CYP90B27, CYP94N1, CYP90G1, and GABAT1 catalyze the first six steps of steroid alkaloid biosynthesis in *V. californicum* to form verazine, a proposed intermediate of the antineoplastic cyclopamine. We refactored the pathway to verazine in *S. frugiperda* Sf9 cells by infections with viruses containing each gene, thereby demonstrating

gene-product function. Verazine exhibits strong antifungal properties and can serve as a synthon for the synthesis of bioactive molecules with additional pharmaceutical properties including liver protection and anti-tumor activities (Gan *et al.* 1993, Jiang *et al.* 2005, Kusano *et al.* 1987, Zhou 2003). We also demonstrated that the precursor in this pathway is 22(*R*)-hydroxycholesterol rather than 22(*S*)-hydroxycholesterol, as previously proposed (Kaneko *et al.* 1977). Our data suggests that dormantinone is not an intermediate in the formation of verazine, and the amino donor for GABAT1 is GABA rather than L-arginine or L-glutamine (Kaneko *et al.* 1976).

The candidate gene selection criteria with emphasis on clades that contained only *Veratrum* genes fitting the Haystack model increased our rate of gene identification. Of the 3,219 genes fitting the Haystack model, 9 cytochrome P450 genes were assigned high priority for biochemical characterization. Seven were tested; 3 catalyzed the first reactions in the verazine biosynthesis pathway. Likewise, 3 GABATs were prioritized for functional characterization; 1 GABAT was found to participate in the verazine pathway. Based upon the success of this methodology, we expect continued, facile elucidation of this and other biochemical pathways from non-model systems.

2.7 Experimental procedures

Plant material and RNA extraction

See Method S1 and Table S9.

Liquid chromatography mass spectrometry (LC-MS/MS) method

Samples were first extracted with either ethanol (see Method S2) for plant tissue or ethyl acetate for Sf9 cell suspensions and enzymatic assays. Liquid chromatographic separation was achieved with 10 µl injections on a LC-20AD (Shimadzu) LC system coupled to a 4000 QTRAP (AB Sciex Instruments) for MS/MS analysis. Separation was achieved using a Phenomenex Gemini C-18 NX column (150 X 2.00 mm, 5 µm) with a flow rate of 0.5 ml/min and the following gradient program [solvent A (0.05% formic acid/0.04% ammonium hydroxide (25%) v/v in H_2O; solvent B (0.05% formic acid/0.04% ammonium hydroxide (25%) v/v in 90% acetonitrile]: Solvent B was held at 20% for 2 min, then 2-11 min 20-30% B, 11 – 18 min 30-100% B, 18-22 min 100% B, 22-23 min 100-20% B, and held at 20% B for an additional 5 minutes. Program parameters included a TurboIonSpray ionization source temperature of 500°C and low resolution for Q1 and Q3 done with MRM (Multiple Reaction Monitoring) scans in the positive ion mode. Specific ion fragments and parameters can be found in Table S10. In conjunction,

EMS (Enhanced MS) scan with a mass range of 380 to 425 *m/z*, and EPI (Enhanced Product Ion) scans for 398, 417, and 418 *m/z* were included. Compound identification was determined by comparison of retention time and fragmentation pattern to the authentic standard cyclopamine (Infinity Pharmaceuticals) (where applicable). Quantitation was performed by plotting peak area versus pmol of standard using Analyst 1.5 (Applied Biosystems).

Gas chromatography mass spectrometry (GC-MS) method

Samples were first extracted with either hexane:isopropanol 3:2 followed by hexane only or ethyl acetate. Dried extracts were derivatized with 40 µl Sylon HTP (Sigma) for 1 hour at 90°C prior to injection with a 7683B autosampler onto a 7890A gas chromatograph coupled to a 5975C mass spectrometer inert XL MSD with triple-axis detector (Agilent Technologies). Both full scan and SIM methods were run in the splitless mode with 1 µl injection volume and a flow rate of 1 ml/min with helium as the carrier gas. Separation was performed on a Zebron ZB-5MSi column (Phenomenex) with guardian 5M (30 m x 0.25 mm x 0.25 µm) with 5% Polysilarylene - 95% Polydimethylsiloxane copolymer composition and 106 relative voltage. The initial temperature of 240°C was held for 5 minutes and increased to 300°C at a rate of 10°C/min and held for 25 minutes. The full scan method measured mass from 50 to 800 amu and ions detected in the SIM mode included: 99.1, 129, 165, 171, 173.1, 187, 261, 314.1, 329.3, 330, 370, 382.3, 417.4, 456.4, 458, 460, 470, 472.3, 486, 546, 560, and 634.

***Veratrum californicum* metabolite extraction for quantitation by LC-MS/MS**

See Method S2.

Transcriptome assembly and determination of relative contig expression

cDNA library construction, Illumina paired-end sequencing, and *de novo* transcriptome assembly were performed at the National Center for Genome Resources (Santa Fe, New Mexico). *Please refer Method S3 for transcriptome assembly details.*

Transcriptome dataset interrogation using Haystack and Plant Tribes

Identification of genes whose expression pattern correlated with accumulation of cyclopamine was determined using the Haystack program (Michael *et al.* 2008, Mockler *et al.* 2007). The LC-MS/MS cyclopamine quantitation data for the different *V. californicum* tissues was used to formulate a model based upon the ratio of biosynthetic tissues. 95 % of the total cyclopamine was found in the subterranean tissues (root, bulb, and rhizome) whereas 5 % was found above ground (leaf, shoot, and flower). For the input model, each subterranean tissue was given a

value of 20 and all above ground tissues including the tissue culture samples was designated 1. Parameters for Haystack were as follows: correlation cut off = 0.7 fold change = 2, p-value = 0.05 and background = 1. Due to the large data input, Haystack analysis was performed on a UNIX server in-house as opposed to the version available online. Annotation data was then merged with the gene outputs from each of the models. Subsequent alignments and phylogenetic analysis were performed using Muscle algorithm (Edgar 2004) and Mega v6.06 (Tamura *et al.* 2011). See Method S4 for dataset interrogation using PlantTribes and a detailed description of methods for contig prioritization.

Construction of viral expression vectors

Candidate contigs obtained from Haystack and phylogenetic analysis were subjected to BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and global alignments to homologous, experimentally characterized gene sequences with the CLC Main Workbench 6.8, for prediction of the open reading frame. Where the reading frame appeared incomplete, Rapid Amplification of cDNA Ends (RACE) was used to obtain the complete coding sequence. *V. californicum* cDNA was prepared from root RNA extracts using M-MLV Reverse Transcriptase (Invitrogen) according to manufacturer's instructions. All primer sequences and PCR programs can be found in Table S11 and S12, respectively. *Please refer to Method S5 for specific cloning details.*

Virus co-transfection, amplification, and protein production

See Method S6 and Table S13.

Extraction of multiple infections for Sf9 *in vivo* product production

See Method S7.

Enzyme assays

Each cytochrome P450 co-expressed with CPR in *S. frugiperda* Sf9 cells and GABAT1 was subjected to individual enzyme assays with the compounds designated in Table S5 including cholesterol (Sigma Aldrich), 22(*R*)-hydroxycholesterol (Sigma Aldrich), 26-hydroxycholesterol [27(25R)-hydroxycholesterol] (Avanti Polar Lipids Inc.), 22(*S*)-hydroxycholesterol (American Radiolabeled Chemicals), 24(*S*)-hydroxycholesterol (American Radiolabeled Chemicals), 4β-hydroxycholesterol (Research Plus Inc.), 7β-hydroxycholesterol (Sigma Aldrich), campesterol (Avanti Polar Lipids Inc.), β-sitosterol (Sigma Aldrich), stigmasterol (Sigma Aldrich), GABA (Sigma Aldrich), L-arginine (Sigma Aldrich), and L-glutamine (Sigma Aldrich) to determine substrate specificity. Compounds were prepared to 1 mM stock solutions of 100% DMSO and diluted with H_2O, except for the amino acids, which were prepared as 200 mM stocks in pure

H_2O. For GC-MS analysis, 5 individual assays per substrate were pooled after incubation at 30°C for 2 hours; one assay produced sufficient product for analysis by LC-MS/MS. Assay conditions were as follows: 80 µl *S. frugiperda* Sf9 cell suspension (obtained by re-suspension of 50 ml viral infected culture pellet in 3.5 ml of 100 mM tricine pH 7.4/ 5 mM thioglycolic acid), 60 mM potassium phosphate buffer pH 8, 1.25 mM NADPH, 7.5 µM substrate, and H_2O in a total volume of 200 µl. Controls were performed with no enzyme and *S. frugiperda* Sf9 cells expressing an unrelated cytochrome P450 (CYP719A14 cheilanthifoline synthase from *Argemone mexicana*), or CPR-only, for each assay.

The initial GABAT1 enzyme assay contained 55 µl *S. frugiperda* Sf9 cell suspension infected with CYP90B27, CYP94N1, CYP90G1, and CPR modified baculoviruses (to provide 22-hydroxycholesterol-26-al substrate), 40 µl *S. frugiperda* Sf9 cells expressing GABAT1, 60 mM potassium phosphate buffer pH 8, 1.5 mM DTT, 100 µM pyridoxal-5-phosphate (PLP), 16 mM GABA, 500 µM NADPH, and H_2O to a total volume of 200 µl. Assay mixes lacking either enzyme or GABA, and control cytochrome P450 assays were run in parallel and each was allowed to proceed for 2 hours at 30°C. Samples were extracted twice with 400 µl ethyl acetate. For determination of the amino group donor, assays were prepared as above for GABAT1 with the following exceptions. The GABAT1 gene was cloned into the EcoRI/NdeI sites of pET28a and expressed in *E.coli* PlusE cells. The GABAT1 protein was purified in the presence of 2 µg/ml PLP using TALON metal affinity resin (Clontech) and PD-10 desalting columns (GE Healthcare) according to manufacturer's instructions to obtain purified enzyme. The substrate 22-hydroxycholesterol-26-al was obtained by extracting 2 ml of Sf9 cells expressing CYP90B27, CYP94N1, and CPR with 2 volumes of ethyl acetate and dried with N_2. The dried extract was resuspended in 400 µl of 10% DMSO. Assays were performed in duplicate and each run with 22-hydroxycholesterol-26-al and either GABA, L-arginine, or L-glutamine alongside CPR only controls and controls using GABAT1 expressed in Sf9 cells. Samples were then dried under N_2, re-suspended in 50-100 µl 80% methanol, and injected onto LC-MS/MS with conditions described above. All cytochrome P450 enzyme assays utilized crude *S. frugiperda* Sf9 protein extracts that contain endogenous metabolites, including cholesterol.

Assays to clarify order of enzymatic transformations

See Method S8 and Figure S11.

Enzymatic product purification for NMR and High Resolution MS for structure elucidation

See Method S9 and Data S2.

Dimedone aldehyde trapping

Enzyme assays containing CYP94N1 and 22(*R*)-hydroxycholesterol as substrate, or CYP90B27 + CYP94N1 utilizing endogenous cholesterol in *S. frugiperda* Sf9 cells as substrate, or CYP90B27 + CYP94N1 and GABAT1, also utilizing endogenous cholesterol in *S. frugiperda* Sf9 cells as substrate with either 80 µl 10 mg/ml dimedone in 10% DMSO or 80 µl 10% DMSO were incubated overnight at 30°C. Assays were extracted twice with 2 volumes ethyl acetate and analyzed by LC-MS/MS. All cytochrome P450 enzymes were co-expressed with CPR.

Sodium borohydride reduction

2 ml *S. frugiperda* Sf9 cells expressing CYP90B27 + CYP94N1 + GABAT1 + CYP90G1 + CPR were extracted twice with equal volume ethyl acetate. Extracts were divided equally, dried under N_2, and re-suspended in 50 µl 80% methanol each. One sample was treated with 50 µl 1 M $NaBH_4$ in 1 M NaOH for 15 minutes. 100 µl H_2O were added to both samples, and each extracted twice with equal volumes of chloroform. Samples were dried under N_2, re-suspended in 50 µl 80% methanol and analyzed by LC-MS/MS as described above. *S. frugiperda* Sf9 cells expressing CPR only were run in parallel as control**.**

Phylogenetic analysis of cytochrome P450 enzymes across species using deep transcriptome sequence data from 1KP and MonAToL projects

See Method S10 and Table S8.

2.8 Acknowledgments

We greatly thank Infinity Pharmaceuticals for providing authentic cyclopamine. This work was supported by funds from Infinity Pharmaceuticals to T.M.K., NSF DEB-1442071 to E.A. Kellogg, NSF DBI-0521250 to PMSF, an Indo-US (IUSSTF) Research Fellowship to A.K.S. and NIH 1R01DA025197-02 to T.M.K. Analysis performed on the 4000 QTRAP and LTQ-Velos Pro Orbitrap was supported by The National Science Foundation under Grants DBI-0521250 and DBI-0922879, respectively. We also thank the National Science Foundation for supporting this research through Assembling the Tree of Life (DEB 0829868), and the 1000 Plants (1KP) initiative, led by GKSW, which was funded by the Alberta Ministry of Innovation and Advanced Education, Alberta Innovates Technology Futures (AITF) Innovates Centres of Research Excellence (iCORE), Musea Ventures, and BGI-Shenzhen.

2.9 Supporting information

2.9.1 Supporting figures

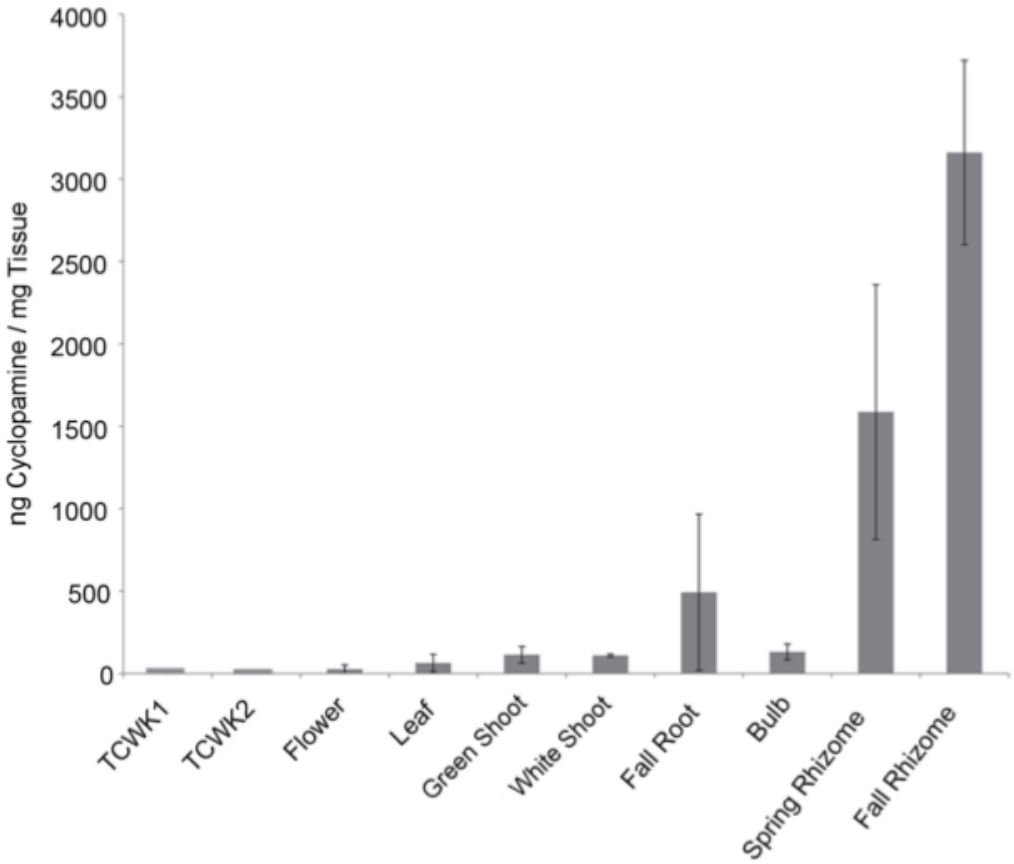

Figure S1. Cyclopamine accumulation profile in *Veratrum californicum.* Each *V. californicum* tissue underwent three independent ethanol extractions followed by LC-MS/MS analysis on a 4000 QTRAP. Quantitation with authentic cyclopamine was accomplished using a standard curve with peak areas. Each value represents ng of alkaloid per mg of tissue; error bars represent standard deviation. Sample dilutions were as follows: 10-fold for tissue culture samples, 1,000-fold for flower, 5,000-fold for leaf and shoot, and 10,000-fold for root, bulb, and rhizome. TCWK1 and TCWK2 stand for tissue culture one- and two weeks after transfer to new media, respectively.

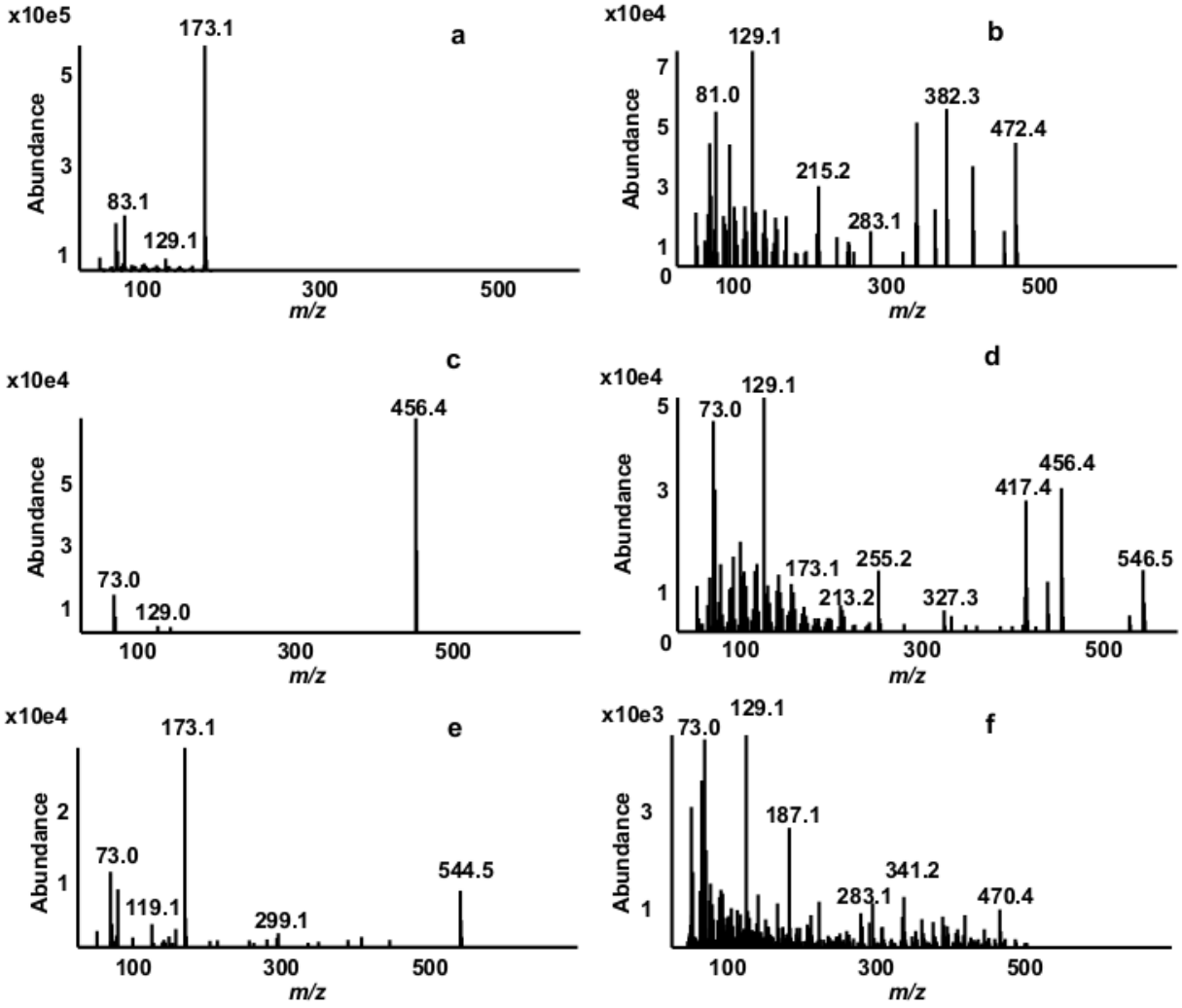

Figure S2. Mass spectra of select derivatized standards and enzymatically formed products. Enzyme assays using recombinant *Veratrum californicum* genes and authentic standards were first extracted with hexane and derivatized with Sylon HTP before GC-MS analysis. (a) Spectrum of 22(*R*)-hydroxycholesterol produced by enzyme assay using *S. frugiperda* Sf9 cells expressing CYP90B27 (cholesterol 22-hydroxylase) and *E. californica* cytochrome P450 reductase (CPR) with pure cholesterol as substrate. (b) Spectrum of 22-keto-cholesterol produced by enzyme assay using Sf9 cells expressing CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase) with pure 22(*R*)-hydroxycholesterol as substrate. (c) Spectrum of pure 7β-hydroxycholesterol. (d) Spectrum of pure 26-hydroxycholesterol. (e) Spectrum of 7β,22-dihydroxycholesterol produced by enzyme assay using Sf9 cells expressing CYP90B27 and CPR with pure 7β-hydroxycholesterol as substrate. (f) Spectrum of 22-keto-26-hydroxycholesterol produce by infection of Sf9 cells with CYP90B27, CYP90G1, CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), and CPR. Mass spectra after background subtraction is shown.

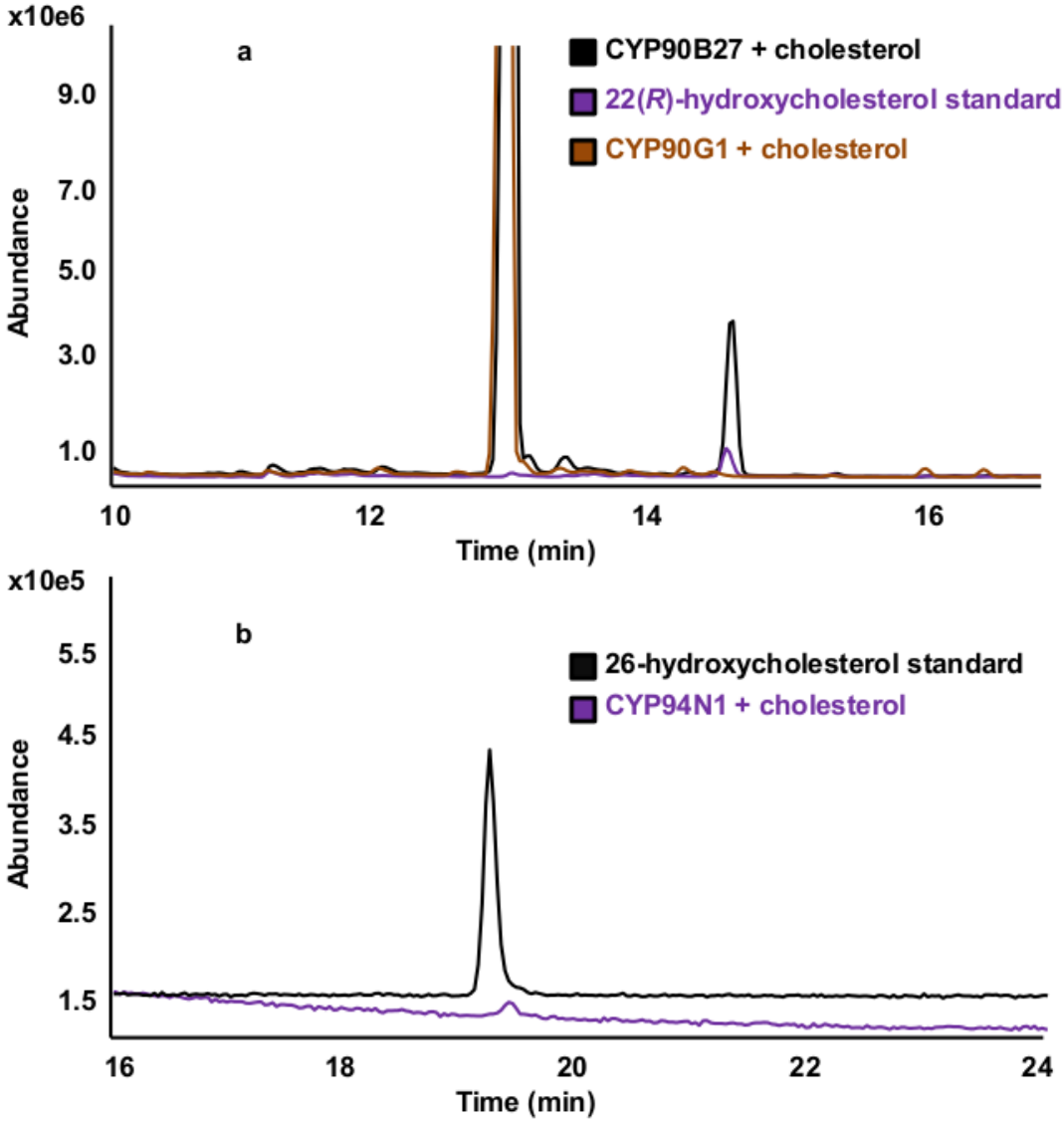

Figure S3. GC-MS analysis of selected *Veratrum californicum* cytochrome P450 enzyme assays with cholesterol. Enzyme assays using *S. frugiperda* Sf9 cells expressing each enzyme were extracted and derivatized with Sylon HTP before GC-MS analysis. (a) Overlay of enzyme assays performed with either CYP90B27 (cholesterol 22-hydroxylase) or CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase) using cholesterol as substrate. Both cytochrome P450 enzymes were co-expressed with *E. californica* cytochrome P450 reductase (CPR). 22(*R*)-Hydroxycholesterol pure standard was included for reference. (b) Enzyme assay of CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase) co-expressed with CPR and cholesterol as substrate, overlaid with pure 26-hydroxycholesterol [27(25R)-hydroxycholesterol] as reference.

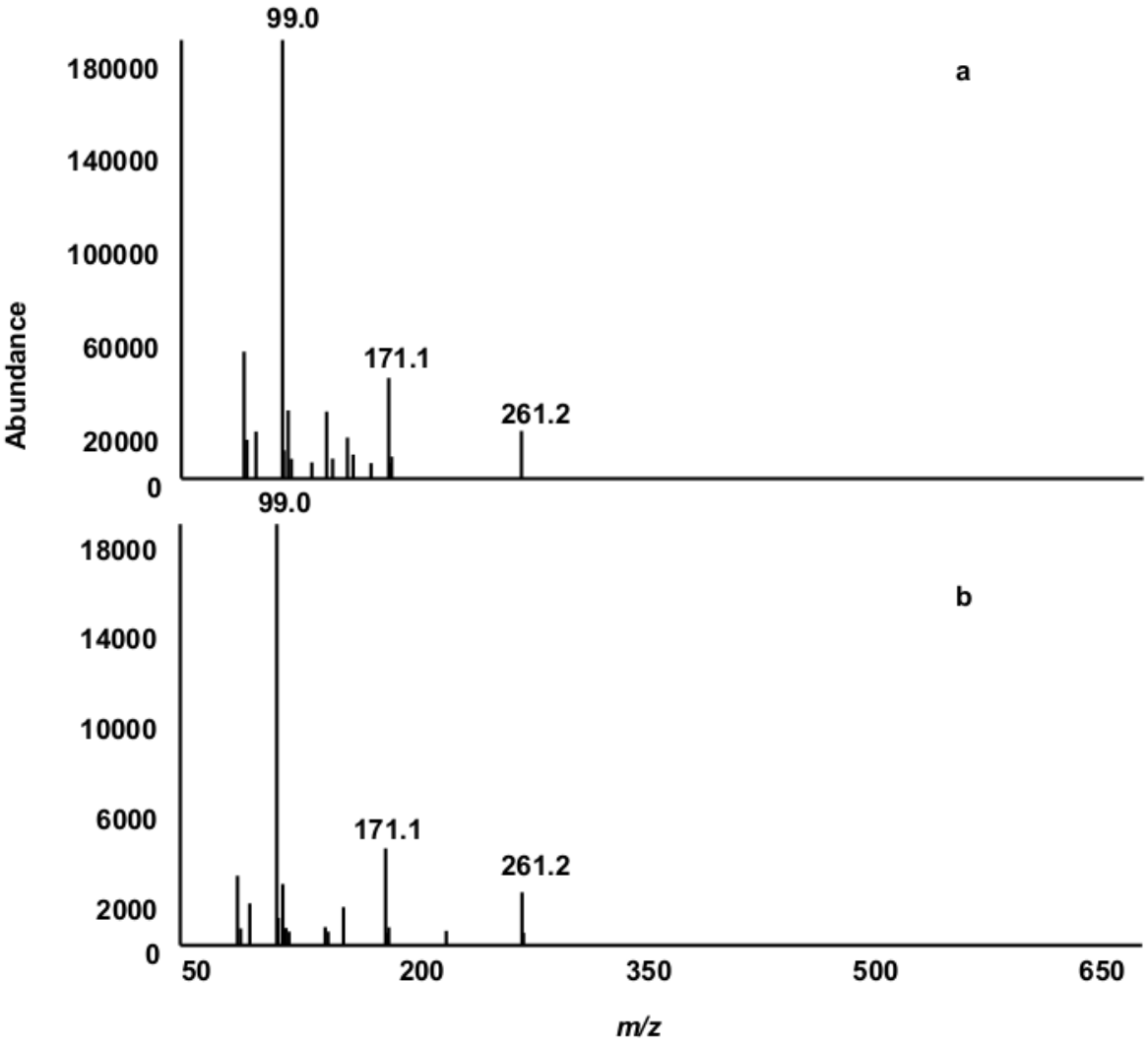

Figure S4. Mass spectra of enzymatically formed 22,26-dihydroxycholesterol. Mass spectrum of product formed by enzyme assays using (a) *S. frugiperda* Sf9 cells expressing *Veratrum californicum* CYP90B27 (cholesterol 22-hydroxylase) and *E. californica* cytochrome P450 reductase (CPR) with pure 26-hydroxycholesterol [27(25*R*)-hydroxycholesterol] as substrate and (b) CYP90B27, *V. californicum* CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), and CPR with pure cholesterol as substrate were extracted and derivatized with Sylon HTP before GC-MS analysis. Mass spectra after background subtraction is shown.

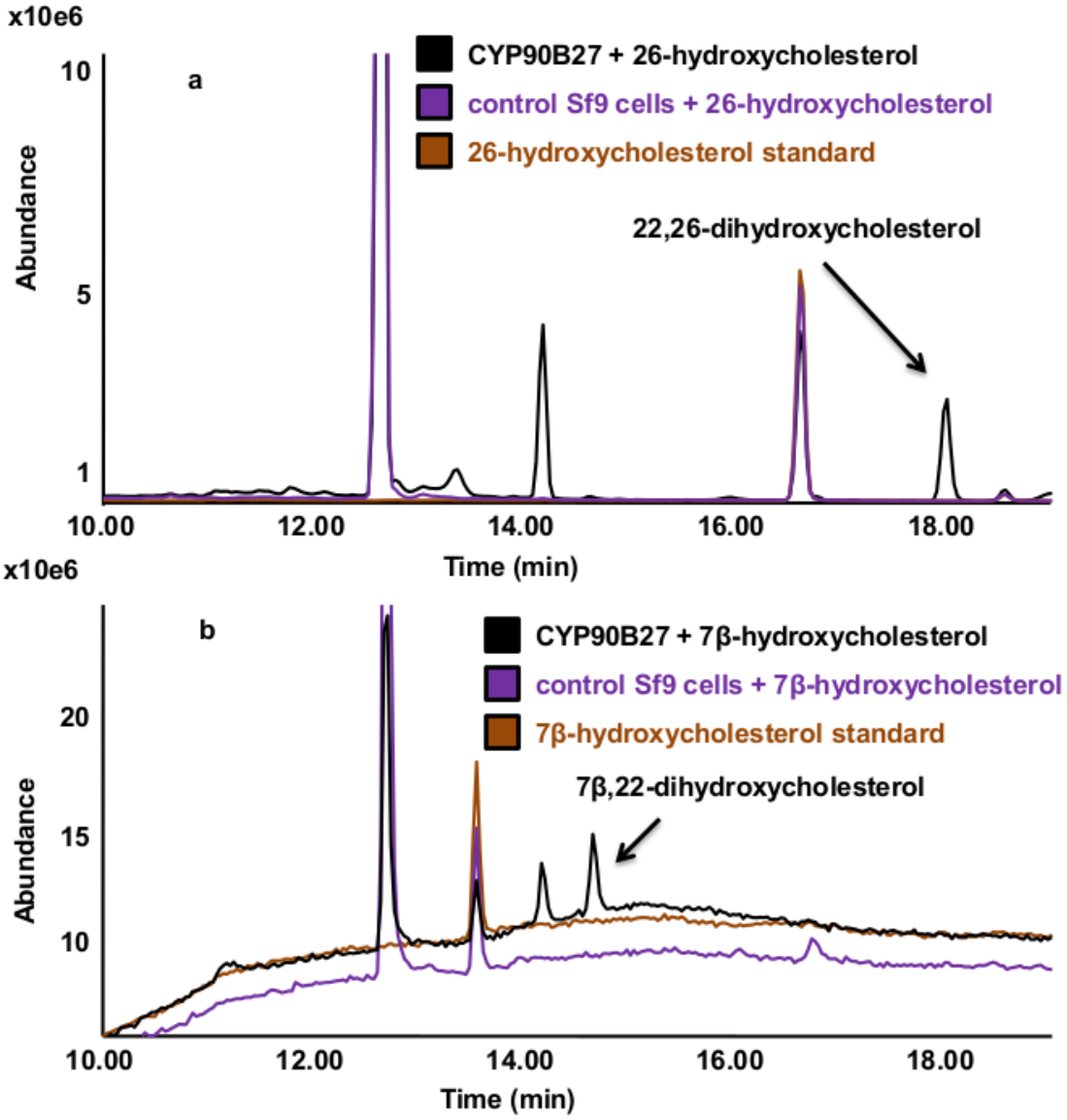

Figure S5. GC-MS analysis of *Veratrum californicum* CYP90B27 with 26-hydroxycholesterol and 7β-hydroxycholesterol. Enzyme assays using *S. frugiperda* Sf9 cells expressing CYP90B27 (cholesterol 22-hydroxylase) and *E. californica* cytochrome P450 reductase (CPR) were extracted and derivatized with Sylon HTP before GC-MS analysis. (a) Enzyme assays using 26-hydroxycholesterol [27(25R)-hydroxycholesterol] as substrate. 26-Hydroxycholesterol [27(25R)-hydroxycholesterol] pure standard was included for reference. Control Sf9 cells expressing CYP719A14, cheilanthifoline synthase from *Argemone mexicana* (an unrelated cytochrome P450), and CPR, were run in parallel with 26-hydroxycholesterol [27(25R)-hydroxycholesterol] as a control. (b) Enzyme assays with 7β-hydroxycholesterol as substrate. 7β-Hydroxycholesterol pure standard was included for reference and Sf9 cells expressing CYP719A14 and CPR were run in parallel with 7β -hydroxycholesterol as a control.

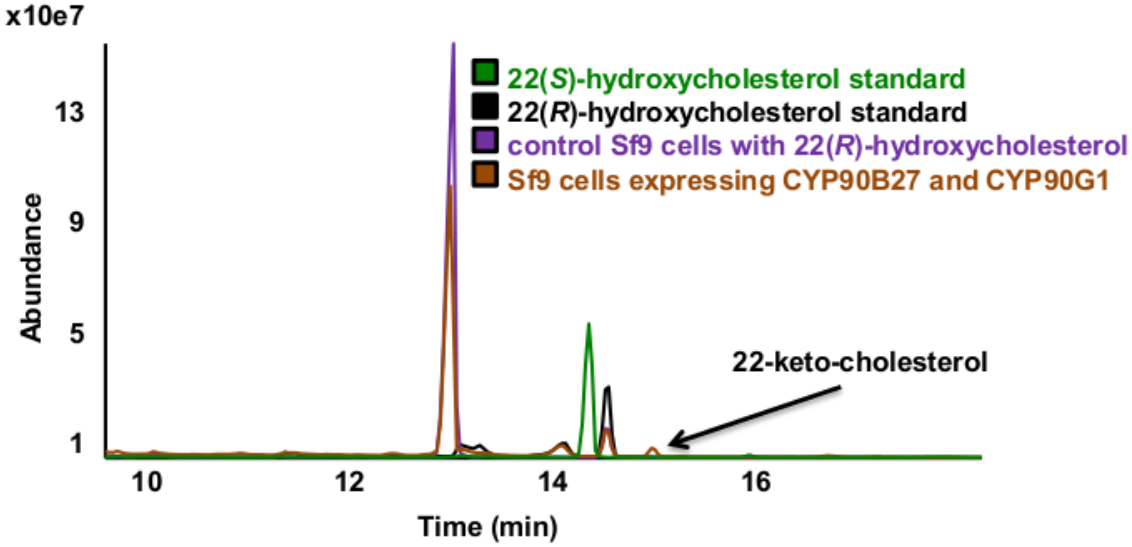

Figure S6. GC-MS analysis of *Veratrum californicum* cytochrome P450 enzymes CYP90B27 and CYP90G1. *S. frugiperda* Sf9 cells expressing enzyme CYP90B27 (cholesterol 22-hydroxylase), CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), and *E. californica* cytochrome P450 reductase (CPR) were extracted and derivatized with Sylon HTP before GC-MS analysis. 22(*R*)-Hydroxycholesterol and 22(*S*)-hydroxycholesterol pure standards were included for reference. Control cells expressing CYP719A14, cheilanthifoline synthase from *A. mexicana* (an unrelated cytochrome P450), and CPR were also assayed with 22(*R*)-hydroxycholesterol as a control.

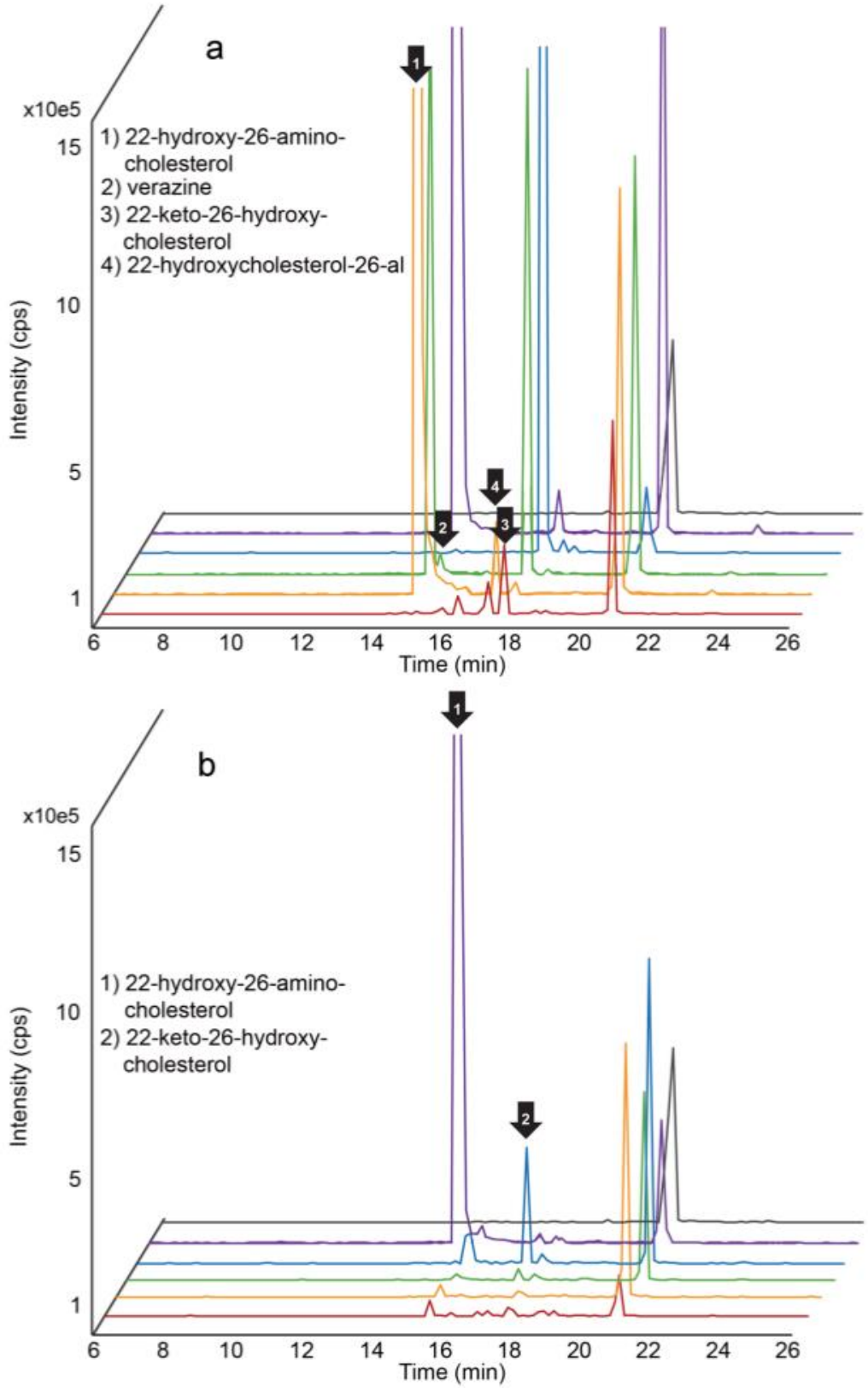
a
x10e5
15
10
5
1
Intensity (cps)
1) 22-hydroxy-26-amino-cholesterol
2) verazine
3) 22-keto-26-hydroxy-cholesterol
4) 22-hydroxycholesterol-26-al
1
2
3
4
6
8
10
12
14
16
18
20
22
24
26
Time (min)
b
x10e5
15
10
5
1
Intensity (cps)
1) 22-hydroxy-26-amino-cholesterol
2) 22-keto-26-hydroxy-cholesterol
1
2
6
8
10
12
14
16
18
20
22
24
26
Time (min)

Figure S7. Determination of the order of the enzymatic steps to verazine. Enzyme assays were performed as described in Figure S11 b using crude *S. frugiperda* Sf9 cells infected with baculovirus containing each *Veratrum californicum* gene. All cytochromes P450 were co-expressed with *E. californica* cytochrome P450 reductase (CPR); GABAT1 (22-hydroxycholesterol-26-al transaminase) was expressed alone. Cholesterol substrate for each initial reaction was provided by Sf9 cells. Extracts were taken at each step and analyzed by LC-MS/MS. Chromatograms for each sample are a combination of product ions for clarity. (a) Chromatograms of assays run initially with CYP90B27 (cholesterol 22-hydroxylase) combined with CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase). Each colored chromatograph corresponds to the following: **Red**-CYP90B27 + CYP94N1, **Orange**-CYP90B27 + CYP94N1, extracted, then added as substrate to assay with GABAT1, **Green**-CYP90B27 + CYP94N1, extracted, then added as substrate to assay with GABAT1, extracted, and added as substrate to CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), **Blue**-CYP90B27 + CYP94N1 extracted and added as substrate to CYP90G1, **Purple**-CYP90B27 + CYP94N1, extracted, and added as substrate to CYP90G1, extracted, and added as substrate to GABAT1, **Grey**-CPR only control. (b) Chromatograms of assays run initially with CYP90B27 + CYP90G1. Each colored chromatogram corresponds to the following: **Red**-CYP90B27 + CYP90G1, **Orange**-CYP90B27 + CYP90G1, extracted, and added as substrate to GABAT1, **Green**-CYP90B27 + CYP90G1, extracted, and added as substrate to GABAT1, extracted, then added as substrate to CYP94N1, **Blue**-CYP90B27 + CYP90G1, extracted, and added as substrate to CYP94N1, **Purple**-CYP90B27 + CYP90G1, extracted, and added as substrate to CYP94N1, extracted, then added as substrate to GABAT1, **Grey**-CPR only control.

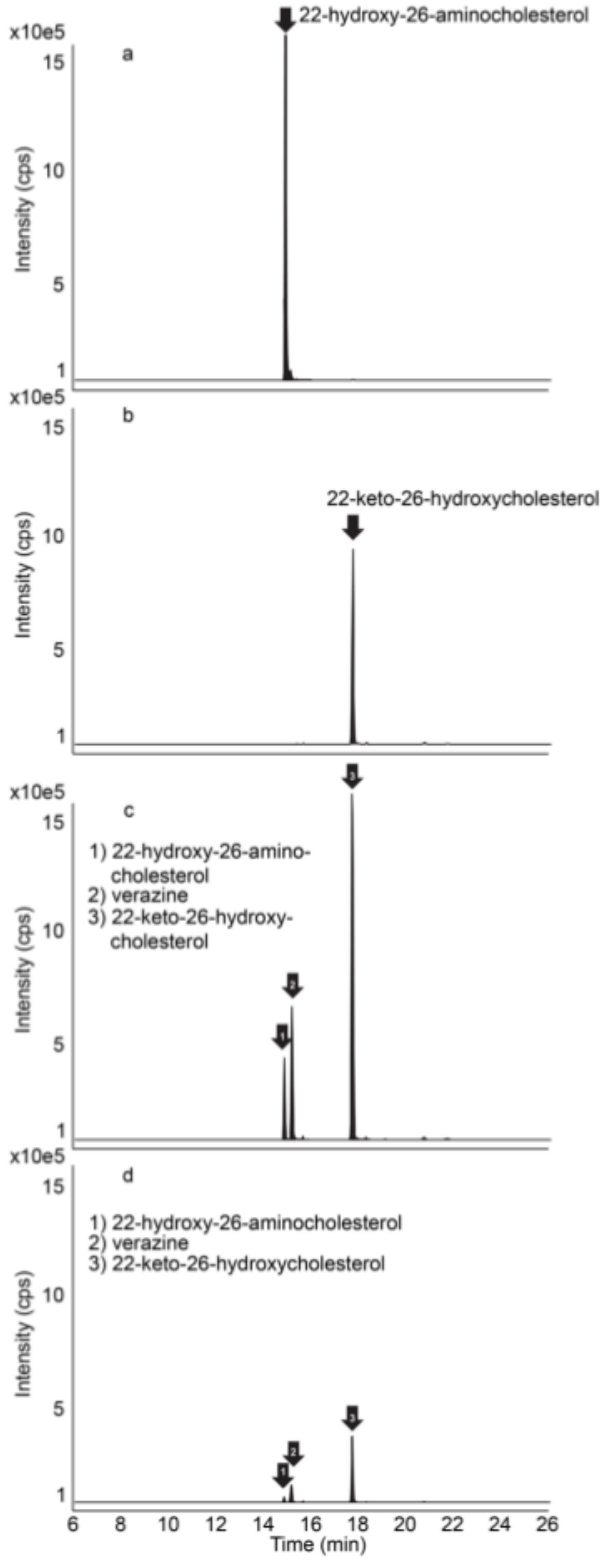

Figure S8. LC-MS/MS of *S. frugiperda* Sf9 extracts expressing *Veratrum californicum* genes. Sf9 cells infected with several combinations of baculovirus containing select genes from *V. californicum* were extracted and analyzed by LC-MS/MS. (a) Extract of Sf9 infection with CYP90B27 (cholesterol 22-hydroxylase), CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), GABAT1 (22-hydroxycholesterol-26-al transaminase), and CPR, (b) extract of Sf9 infection with CYP90B27, CYP94N1, CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), and CPR, (c) extract of Sf9 infection with CYP90B27, CYP94N1, GABAT1, CYP90G1, and CPR, (d) extract of Sf9 infection with CYP90B27, CYP94N1, CYP90G1, *S. lycopersicum* GABA transaminase isozyme 2, and CPR. MRM signals for each metabolite were combined and shaded for clarity. CPR refers to *E. californica* cytochrome P450 reductase.

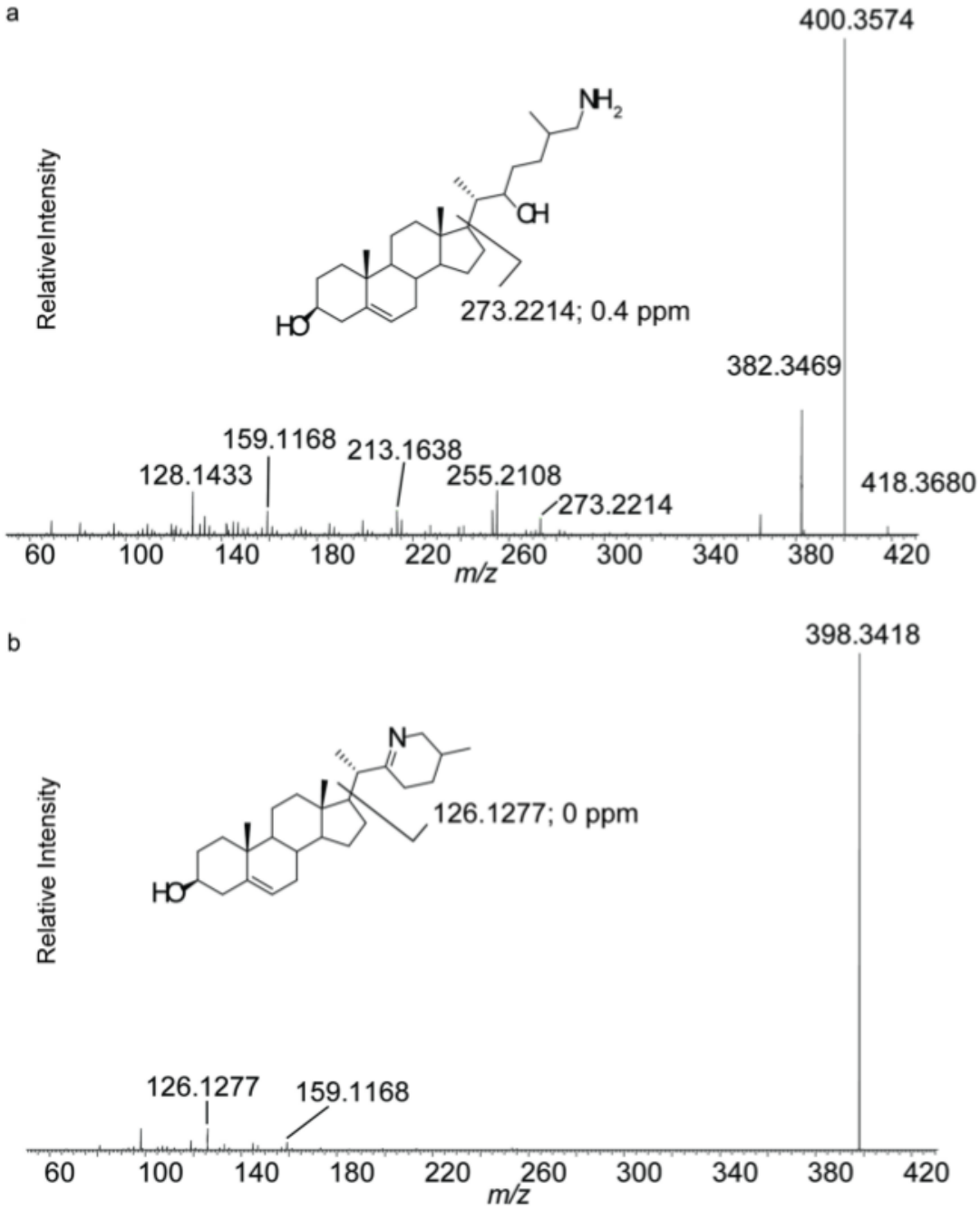

Figure S9. High-resolution mass spectrometric analysis of enzymatically formed 22-hydroxy-26-aminocholesterol and verazine. Purified (a) 22-hydroxy-26-aminocholesterol and (b) verazine were analyzed by high-resolution mass spectrometry with a LTQ-Orbitrap Velos Pro by direct infusion for structural identification. Key fragments for each are indicated.

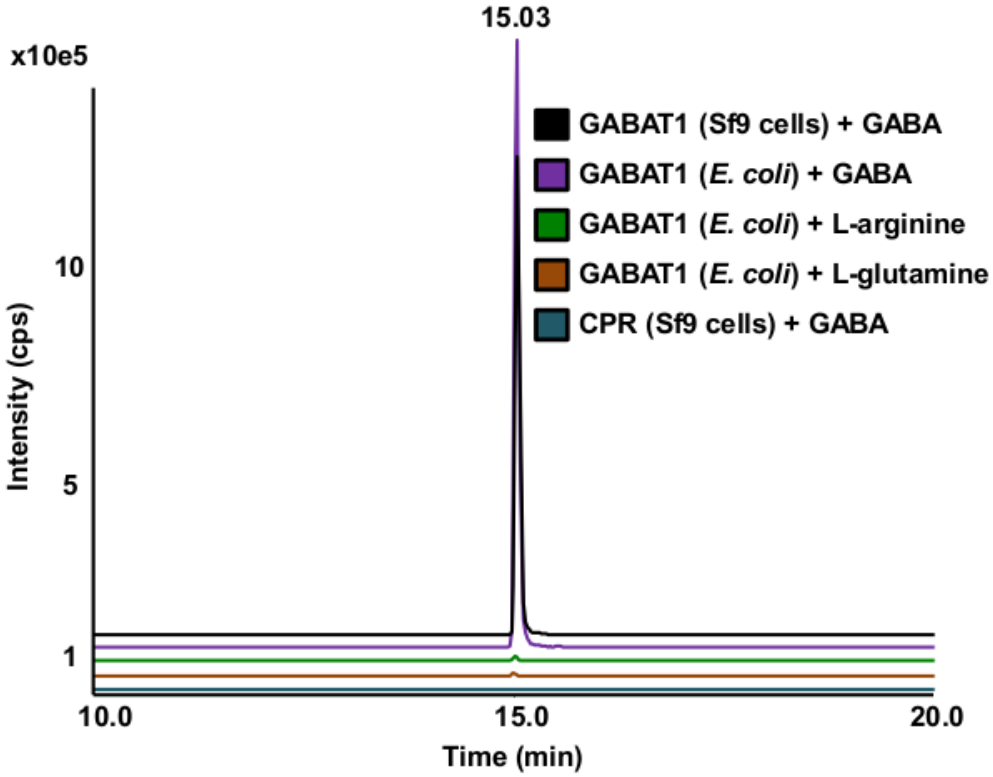

Figure S10. LC-MS/MS of enzyme assays with his-tag purified GABAT1. Recombinant GABAT1 was his-tag purified from *E.coli* PLUS E cells and used in enzyme assays with the substrate 22-hydroxycholesterol-26-al and GABA, L-arginine, or L-glutamine to determine the amino group donor. Sf9 cells infected with CPR were used as a negative control, and Sf9 cells infected with GABAT1 were used as a positive control. MRM signal 418/400 is presented for each assay. CPR refers to *E. californica* cytochrome P450 reductase. Each assay was performed in duplicate, with one representative chromatogram shown.

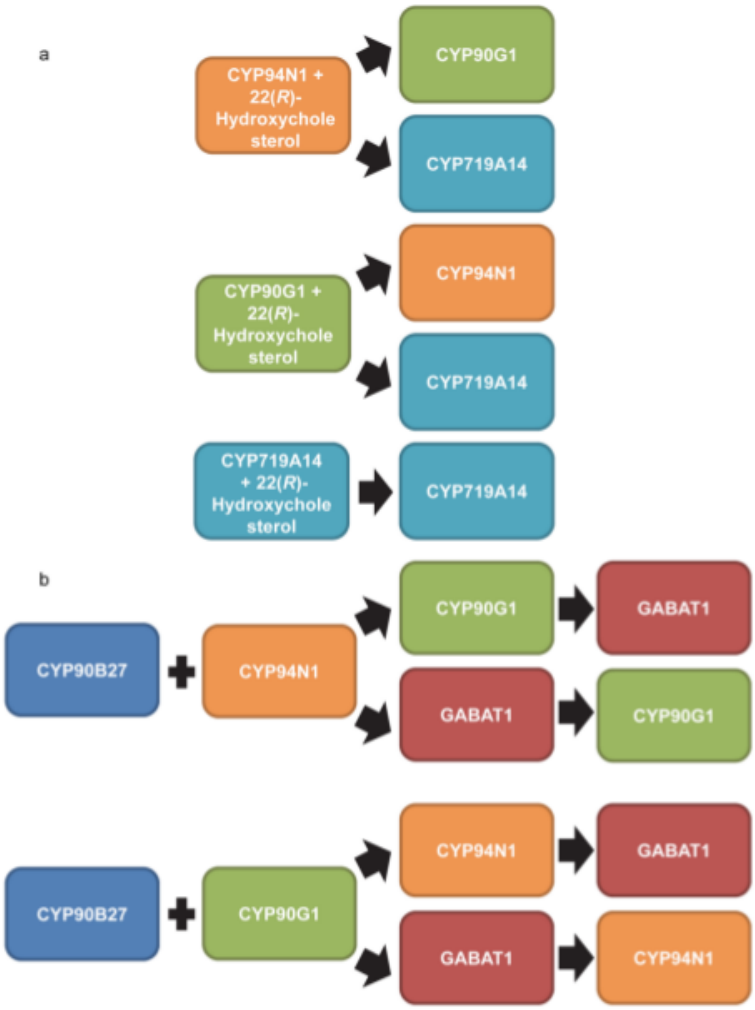

Figure S11. Enzyme assay workflow for clarification of cyclopamine biosynthetic pathway. Assays were performed with crude *S. frugiperda* Sf9 cells infected with baculovirus containing selected *Veratrum californicum* genes. Each arrow represents an extraction step; the resulting product was utilized as substrate for the subsequent enzyme assay. All cytochromes P450 were co-expressed with *E. californica* cytochrome P450 reductase (CPR). (a) 12 assays each of CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), and CYP719A14 (control cytochrome P450) each with pure 22(*R*)-hydroxycholesterol were incubated and extracted. Dried extracts from each were split and used as substrate in the reactions according to panel a. Final extracts were analyzed by GC-MS with results shown in Figure S12. (b) 8 assays of CYP90B27 (cholesterol 22-hydroxylase) and CYP94N1 and 8 assays of CYP90B27 and CYP90G1 were incubated and extracted. Dried extracts were split and used as substrate in 2 more enzyme assays, 4 reactions each as indicated by panel b. These were extracted then split into 2 more assays, 2 reactions each in reactions indicated by panel b. GABAT1 refers to 22-hydroxycholesterol-26-al transaminase. Products at each extraction step were analyzed by LC-MS/MS with results shown in Figure S7.

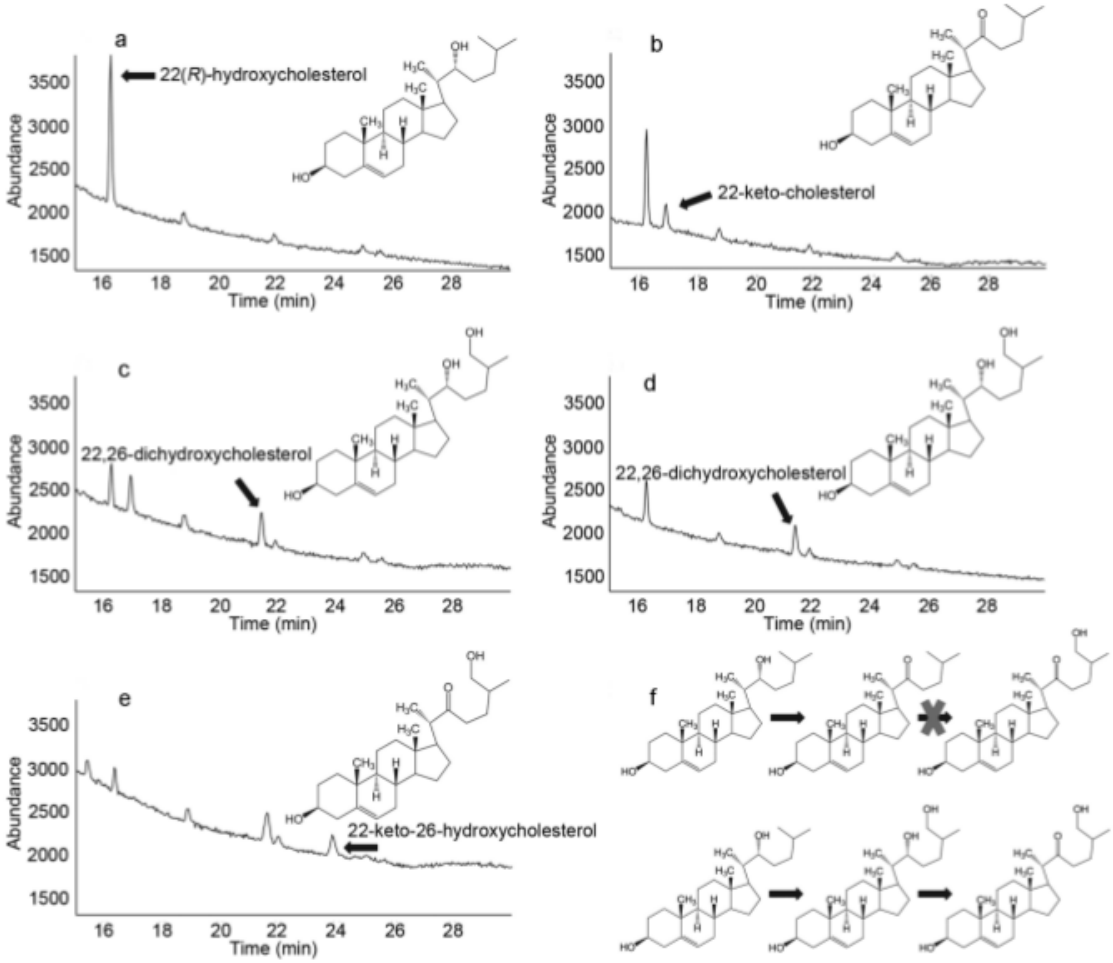

Figure S12. Enzyme assays for biosynthetic pathway order clarification in *Veratrum californicum* using GC-MS. Assays were completed according to Figure S11 a and analyzed by GC-MS in SIM mode. (a) Assay with an unrelated cytochrome P450 as control (CYP719A14, cheilanthifoline synthase from *A. mexicana)* co-expressed with *E. californica* cytochrome P450 reductase (CPR) and authentic 22(*R*)-hydroxycholesterol was extracted, dried, and used as substrate in an assay with the same control cytochrome P450 + CPR. (b) Assay with CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase) + CPR and 22(*R*)-hydroxycholesterol was extracted and used as substrate in assay with control cytochrome P450 + CPR. (c) Assay with CYP90G1 + CPR and 22(*R*)-hydroxycholesterol was extracted and used as substrate in assay with CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase) + CPR. (d) Assay with CYP94N1 + CPR was extracted and used as substrate in assay with control cytochrome P450 + CPR. (e) Assay with CYP94N1 + CPR and 22(*R*)-hydroxycholesterol was extracted and used as substrate in assay with CYP90G1 + CPR. (f) Schematic of possible transformation order; X indicating which reaction was not observed. Ion 187 was extracted for the presented chromatograms.

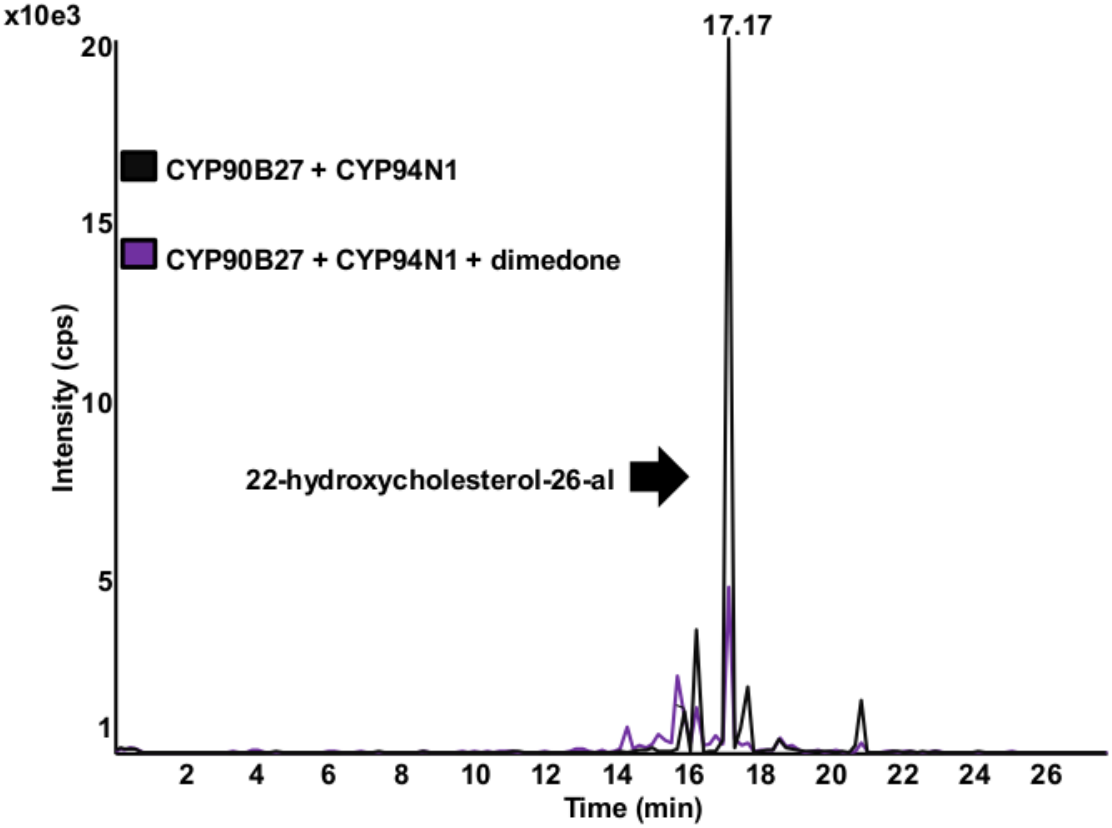

Figure S13. Trapping of enzymatically formed aldehyde intermediate. *S. frugiperda* Sf9 cell extracts expressing *Veratrum californicum* CYP90B27 (cholesterol 22-hydroxylase) co-expressed with *E. californica* cytochrome P450 reductase (CPR) and *V. californicum* CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase) co-expressed with CPR were mixed for enzyme assay using cholesterol provided by the crude cell extract as substrate. Assays performed without dimedone (black) and with dimedone (purple) were analyzed by LC-MS/MS. Chromatograms were obtained by overlay of Enhanced Product Ion scans (EPI) for molecular mass 417.

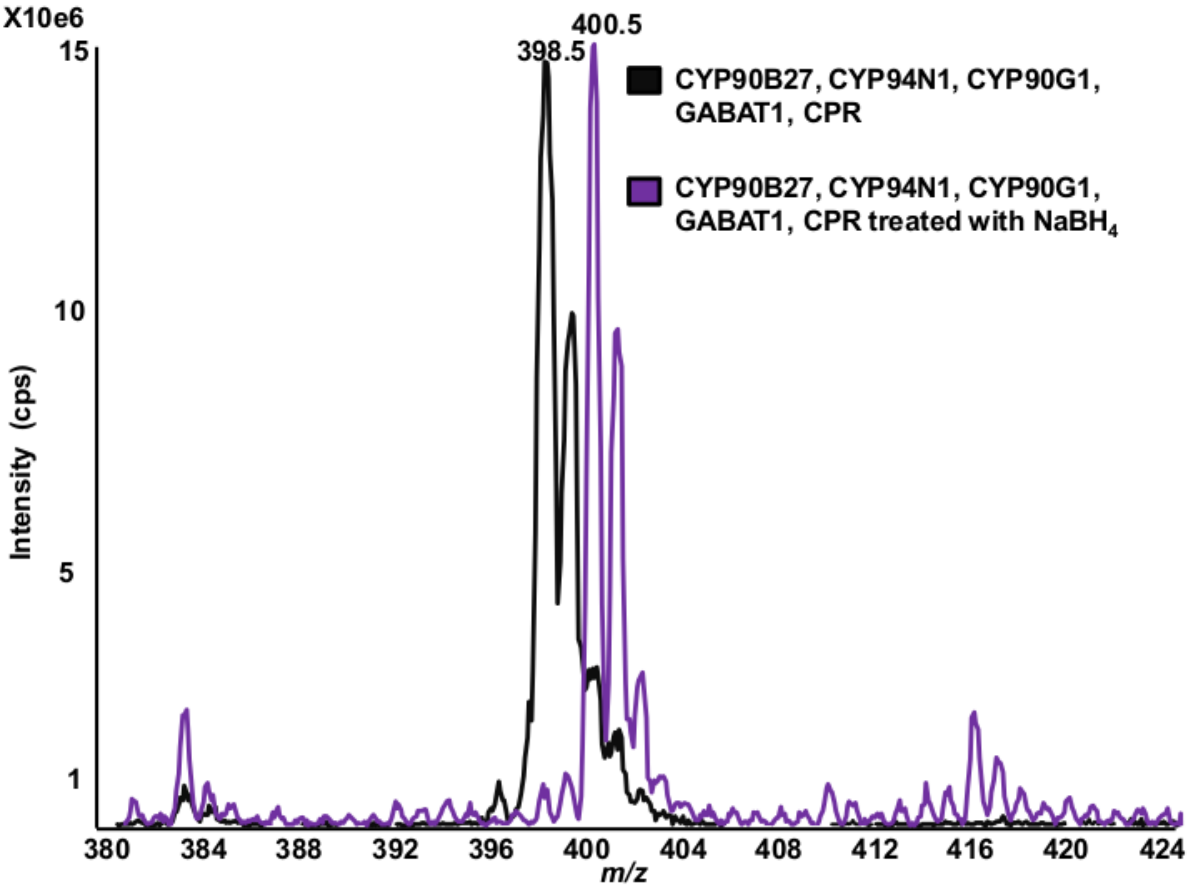

Figure S14. Derivatization of enzymatically formed verazine. *S. frugiperda* Sf9 cells expressing the *Veratrum californicum* genes CYP90B27 (cholesterol 22-hydroxylase), CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), GABAT1 (22-hydroxycholesterol-26-al transaminase), CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), and *E. californica* cytochrome P450 reductase (CPR) were extracted and analyzed by LC-MS/MS either directly (black) or after treatment with $NaBH_4$ (purple). Enhance MS scans detecting ions 380-425 m/z are presented.

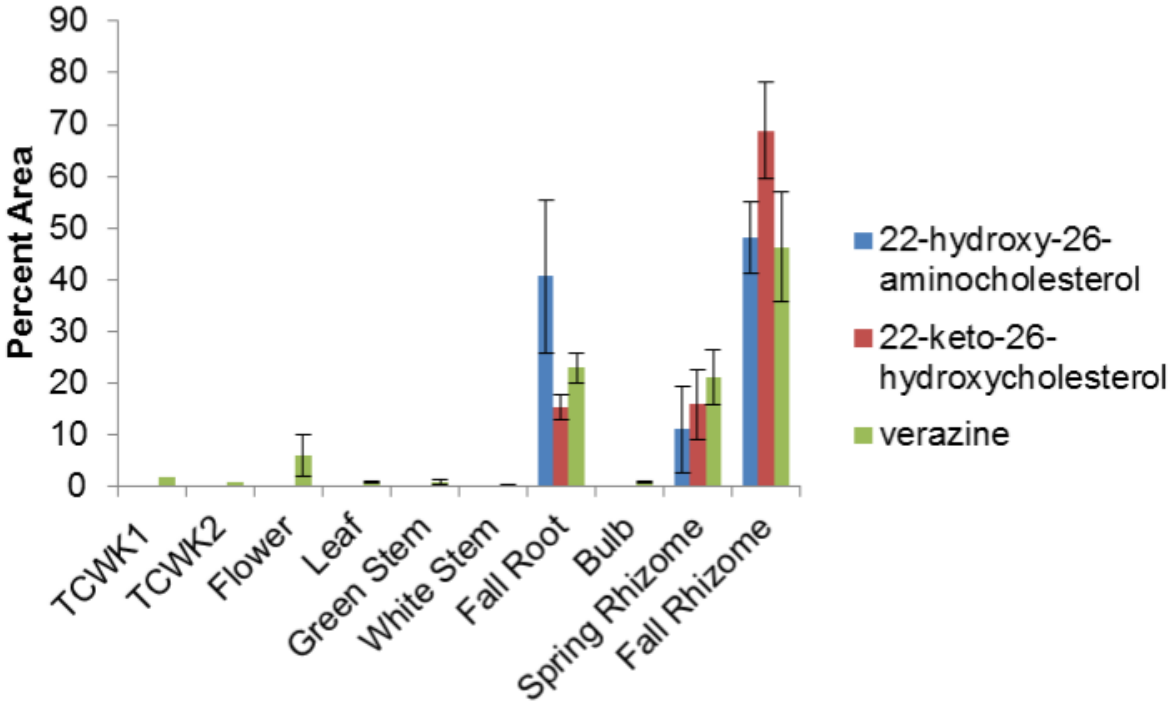

Figure S15. Relative tissue specific accumulation of selected *Veratrum californicum* metabolites. Relative quantities of metabolites 22-hydroxy-26-aminocholesterol, 22-keto-26-hydroxycholesterol, and verazine are shown for each tissue extract by percent of total peak area; error bars show standard deviation for three replicates as determined by LC-MS/MS. TCWK1 and TCWK2 stand for tissue culture one- and two weeks after transfer to new media, respectively.

Figure S16. Phylogenetic tree of cytochrome P450 enzymes using transcriptome data from 1KP and MonAToL sequencing projects. The maximum likelihood tree depicting relationships of P450 genes from across angiosperms and other land plants demonstrates conservation of ancient diversification within gene families. Both ancient and recent gene duplications arise within angiosperms with the P450 gene family. Sequences were obtained by alignment of *Veratrum californicum* P450 enzymes CYP90B27, CYP90G1, and CYP94N1, and the *S. lycopersicum* genes GAME4, GAME 6, GAME 7, and GAME 8 to homologous sequences found in the 1kP and MonAToL sequencing projects. Refer to Table S8 for species key and sequence source.

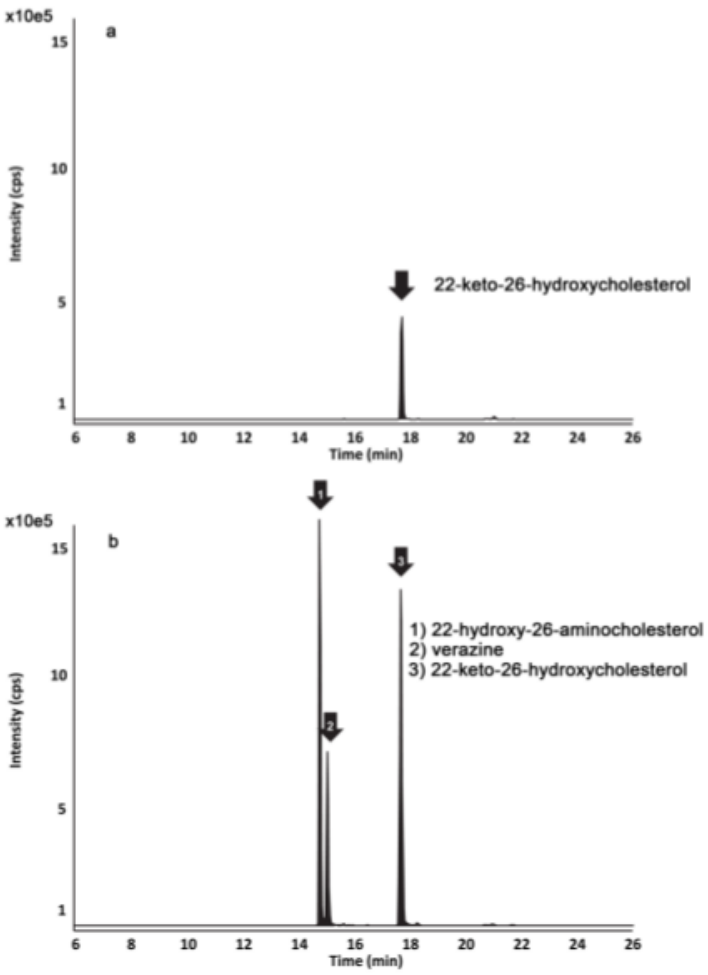

Figure S17. LC-MS/MS analysis of *Veratrum californicum* GABATs. *S. frugiperda* Sf9 cells we co-transformed with *E. californica* cytochrome P450 reductase (CPR), *V. californicum* cytochrome P450 enzymes: CYP90B27 (cholesterol 22-hydroxylase), CYP94N1 (22-hydroxycholesterol 26-hydroxylase/oxidase), CYP90G1 (22-hydroxy-26-aminocholesterol 22-oxidase), and (a) GABAT2 or (b) GABAT1 (22-hydroxycholesterol-26-al transaminase). Extracts were analyzed by LC-MS/MS; ions for each peak were combined and shaded for clarity.

2.9.2 Supporting tables

Table S1. Sequence reads and length generated for each RNA sample (paired end).

Tissue Used for RNA	Sequence Reads Per File	Average Read Length
Bulb	8,870,015	50
Flower	10,499,190	50
Leaf	8,811,375	50
Fall Rhizome	10,558,614	50
Spring Rhizome	11,237,164	50
Fall Roots	8,026,623	50
Green Shoots	10,021,063	50
White Shoots	10,563,408	50
Tissue Culture Week 1	9,946,944	50
Tissue Culture Week 2	12,413,973	50

Table S2. Transcriptome statistics provided by the NCGR.

Sequences	56994
Bases	41106915
Minimum length	100
Maximum length	16273
N50	1471
B1000*	64.4%

B2000**	35.9%

*B1000 is the percent of scaffolds greater than 1000 bases and **B2000 is the percent of scaffolds greater than 2000 bases.

Table S3. Selected top-scoring cytochrome P450 candidate cDNAs for the enzymatic conversion of cholesterol to cyclopamine.

Gene ID	**Putative function**
>medp_verca-20110208|2398	similar to CYP71D unknown function*
>medp_verca-20110208|31930	similar to CYP71D unknown function
>medp_verca-20110208|10041	similar to CYP728 taxane 13a-hydroxylase*
>medp_verca-20110208|13942	similar to CYP734 brassinolide C-26 hydroxylase*
>medp_verca-20110208|13284	similar to CYP90B1 steroid C-22 hydroxylase*+
>medp_verca-20110208|18017	similar to CYP90B1 steroid C-22 hydroxylase
>medp_verca-20110208|18580	similar to CYP90B1 steroid C-22 hydroxylase*+
>medp_verca-20110208|2646	similar to CYP90B1 steroid C-22 hydroxylase*
>medp_verca-20110208|32399	similar to CYP90B1 steroid C-22 hydroxylase
>medp_verca-20110208|12709	similar to CYP94D unknown function*

*indicates contigs that were successfully cloned and tested
+contigs are homologs, and exhibited the same function

Table S4. Selected top-scoring transaminases in the steroid alkaloid biosynthetic pathway to cyclopamine.

Gene ID	**Putative function**
>medp_verca-20110208|12217	aminotransferase ACS10
>medp_verca-20110208|12084	gamma aminobutyrate transaminase 1, mitochondrial-like*

>medp_verca-20110208|5285	1-aminocyclopropane-1-carboxylate synthase
>medp_verca-20110208|28717	aminotransferase ACS12-like*
>medp_verca-20110208|15871	histidinol-phosphate aminotransferase, chloroplastic-like*
>medp_verca-20110208|10159	cysteine desulfurase 1
>medp_verca-20110208|1461	methionine S-methyltransferase

*indicates contigs that were successfully cloned and tested.

Table S5. Substrate testing for cytochrome P450 enzymes co-expressed with CPR. Production of a detectable product is indicated by a (+). CYP90B27 refers to cholesterol 22-hydroxylase, CYP90G1 refers to 22-hydroxy-26-aminocholesterol 22-oxidase, and CYP94N1 refers to 22-hydroxycholesterol 26-hydroxylase/oxidase.

	CYP90B27	CYP90G1	CYP94N1
cholesterol	+	-	-
22(*R*)-hydroxycholesterol	+	+	+
22(*S*)-hydroxycholesterol	-	-	-
26-hydroxycholesterol	+	-	N/A
22,26-dihydroxycholesterol	N/A	+	N/A
22-keto-cholesterol	N/A	N/A	-
4β-hydroxycholesterol	-	-	-
7β-hydroxycholesterol	+	-	-
24(*S*)-hydroxycholesterol	-	-	-
campesterol	-	N/A	N/A
β-sitosterol	-	N/A	N/A
stigmasterol	-	N/A	N/A

Table S6. Assigned enzyme names.

Transcriptome derived contig designations	**CYP designation**	**Accession numbers**	**Assigned name based on function**
>medp_verca-20110208|2646	CYP90B27v1	KJ869252	Cholesterol 22-hydroxylase
	CYP90B27v2	KJ869253	Cholesterol 22-hydroxylase
>medp_verca-20110208|12709	CYP94N1v1	KJ869254	22-Hydroxycholesterol 26-hydroxylase/oxidase
	CYP94N1v2	KJ869255	22-Hydroxycholesterol 26-hydroxylase/oxidase
	CYP94N2v1	KJ869256	22-Hydroxycholesterol 26-hydroxylase/oxidase
	CYP94N2v2	KJ869257	22-Hydroxycholesterol 26-hydroxylase/oxidase
>medp_verca-20110208|12084	N/A	KJ869262	22-Hydroxycholesterol-26-al transaminase
	N/A	KJ869263	22-Hydroxycholesterol-26-al transaminase

	N/A	KJ869264	22-Hydroxycholesterol-26-al transaminase
>medp_verca-20110208|13284	CYP90G1v1	KJ869258	22-Hydroxy-26-aminocholesterol 22-oxidase
	CYP90G1v2	KJ869261	22-Hydroxy-26-aminocholesterol 22-oxidase
	CYP90G1v3	KJ869260	22-Hydroxy-26-aminocholesterol 22-oxidase
	CYP90G2	KJ869259	22-Hydroxy-26-aminocholesterol 22-oxidase

Table S7. Accession numbers and both putative and determined function for sequences in cytochrome P450 phylogenetic tree.

Name On Tree	Species Name	Accession Number	Function
VrCYP90A2	*Vigna radiata*	AF279252	3-Epi-6-deoxocathasterone 23-monooxygenase
AtCYP90A1	*Arabidopsis thaliana*	AY087526.1	C-3 oxidation of the early brassinosteroid intermediates
SlCYP90A5	*Solanum lycopersicum*	XM_004240898	Steroid 22-alpha-hydroxylase
ZeCYP90A11	*Zinnia elegans*	AB231153	Steroid 23-alpha-hydroxylase
OsCYP90A3	*Oryza sativa*	AB206580	Brassinosteroid c-23-hydroxylase
AtCYP90B1	*Arabidopsis*	AF412114	Steroid 22-alpha-hydroxylase

	thaliana		
SlCYP90B1	*Solanum lycopersicum*	Solyc02g085360.2.1	Steroid 22-alpha-hydroxylase
OsCYP90B2	*Oryza sativa*	AB206579.1	Steroid 22-alpha-hydroxylase
AtCYP90D1	*Arabidopsis thaliana*	NM_112223.2	3-Epi-6-deoxocathasterone 23-monooxygenase
AtCYP90C1	*Arabidopsis thaliana*	NM_119801	3-Epi-6-deoxocathasterone 23-monooxygenase
PtCYP90D1	*Populus trichocarpa*	Potri.001G200100.1	3-Epi-6-deoxocathasterone 23-monooxygenase
OsCYP90D2	*Oryza sativa*	NM_001048832.1	Steroid 23-alpha-hydroxylase
SmCYP90F1v1	*Selaginella moellendorffii*	XM_002968284.1	Unknown
SmCYP90E1	*Selaginella moellendorffii*	XM_002979568.1	Unknown
SlCYP724B1	*Solanum lycopersicum*	Solyc02g093540.2.1	Unknown
OsCYP724B1	*Oryza sativa*	NM_001059582.1	Steroid 22-alpha-hydroxylase
CYP90G1 (13284)	*Veratrum californicum*	KJ869260	22-Hydroxy-26-aminocholesterol 22-oxidase
CYP90B27 (2646)	*Veratrum californicum*	KJ869252	Cholesterol 22-hydroxylase
MtCYP718A8	*Medicago truncatula*	XM_003617407	Taxane 13-alpha-hydroxylase
AaCYP716A14	*Artemisia annua*	DQ363134	Taxadiene 5-alpha-hydroxylase
MtCYP716A12	*Medicago truncatula*	DQ335781	3-Epi-6-deoxocathasterone 23-monooxygenase
StCYP716A13	*Solanum tuberosum*	XM_006338067	Unknown
SlCYP88B1 (GAME4)	*Solanum lycopersicum*	XM_004251512.1	Steroid 26-oxidase
ZmCYP88A1	*Zea mays*	NM_001112116.1	Early step in gibberellin biosynthesis

LjCYP88D4	*Lotus japonicus*	AB433177	Beta-amyrin 11-oxidase
MtCYP88D3	*Medicago truncatula*	AB433176.1	Unknown
GuCYP88D6	*Glycyrrhiza uralensis*	AB433179.1	Beta-amyrin 11-oxidase
AtCYP710A1	*Arabidopsis thaliana*	NM_129002.2	C-22 sterol desaturase
VsCYP94A1	*Vicia sativa*	AF030260	Fatty acid omega-hydroxylase
AtCYP94B3	*Arabidopsis thaliana*	BT015841.1	Converts JA-Ile to 12-hydroxy-JA-Ile
PpCYP94C	*Physcomitrella patens*	XM_001763154.1	Unknown
SmCYP94J1	*Selaginella moellendorffii*	XM_002979639.1	Medium chain fatty acid hydroxylase
SlCYP94A1	*Solanum lycopersicum*	Solyc09g066150.1.1	Cytochrome P450-dependent fatty acid hydroxylase
NtCYP94A4	*Nicotiana tabacum*	AF092915	Fatty acid hydroxylase
CYP94N1 (12709)	*Veratrum californicum*	KJ869255	22-Hydroxycholesterol 26-hydroxylase/oxidase
PCYP86A22	*Petunia x hybrida*	DQ099538.1	Fatty acid omega-hydroxylase
AtCYP86A1	*Arabidopsis thaliana*	NM_125276.2	Fatty acid omega-hydroxylase
OsCYP714B1	*Oryza sativa*	NM_001067187.1	Gibberellin 13-oxidase
OsCYP714AD1	*Oryza sativa*	AY987039.1	16-Alpha 17-epoxidation on non-13-hydroxylated gibberellins
SlCYP734A7	*Solanum lycopersicum*	NM_001247011	Castasterone 26-hydroxylase
AtCYP734A1	*Arabidopsis thaliana*	NM_128228	Steroid 26-hydroxylase; 26-hydroxylase for brassinolide and castasterone
OsCYP74A3	*Oryza sativa*	NM_001052842.1	Converts 13-hydroperoxylinolenic acid to

			12,13-epoxylinolenic acid
OsCYP734A2	*Oryza sativa*	AB488666.1	C-26 oxidation of brassinosteroids
SlCYP72A208 (GAME8b)	*Solanum lycopersicum*	NM_001247565.1	22-Hydroxycholesterol 26-hydroxylase
NpCYP72A2	*Nicotiana plumbaginifolia*	U35226.3	Unknown
SlCYP72A188 (GAME6)	*Solanum lycopersicum*	XM_004243590.1	Steroidal alkaloid biosynthesis
CrCYP72A1	*Catharanthus roseus*	L10081	Secologanin synthase
GuCYP72A154	*Glycyrrhiza uralensis*	AB558152	Beta-amyrin oxidase
MtCYP72A59	*Medicago truncatula*	DQ335783.1	Unknown
StCYP72A56	*Solanum tuberosum*	XM_006348844	Unknown
SlCYP72A186 (GAME7)	*Solanum lycopersicum*	XM_004244225.1	Putative cholesterol 22-hydroxylase
NtCYP72A57	*Nicotiana tabacum*	DQ350355.1	Unknown
AtCYP72C1	*Arabidopsis thaliana*	NM_101566	Brassinosteroid inactivation
AsCYP51H10	*Avena strigosa*	DQ680852.1	Beta-amyrin 12,13 epoxidase 16-hydroxylase
MeCYP71E7	*Manihot esculenta*	AY217351	Unknown
TaCYP71F1	*Triticum aestivum*	AB036772	Ferulate 5-hydroxylase
CrCYP71BJ1	*Catharanthus roseus*	HQ901597.1	Tabersonine/lochnericine 19-hydroxylase
LsCYP71BL2	*Lactuca sativa*	HQ439599.1	Costunolide synthase

ZmCYP71C1	*Zea mays*	NM_001175287.1	3-Hydroxyindolin-2-one monooxygenase
PCYP75A1	*Petunia x hybrida*	DQ352142.1	Flavonoid 3',5'-hydroxylase
SlCYP85A1	*Solanum lycopersicum*	NM_001247334.1	6-Deoxocastasterone oxidase
OsCYP76M7	*Oryza sativa*	AK105913.1	Ent-cassadiene c-11α-hydroxylase
SmCYP76AH1	*Salvia miltiorrhiza*	JX422213.1	Ferruginol synthase
SaCYP76F37v1	*Santalum album*	KC533717.1	Bergamotene oxidase
PsCYP720B4	*Picea sitchensis*	HM245403.1	Diterpene oxidation to form the diterpene resin acids isopimaric acid and abietic acid

Table S8. Species key and source for sequences in P450 phylogenetic tree using deep transcriptome sequence data from 1KP and MonAToL projects (Al-Mssallem *et al.* 2013, D'Hont *et al.* 2012).

Species	Source
Zinnia elegans	Genbank (Table S7)
Nicotiana plumbaginifolia	Genbank (Table S7)
Nicotiana tabacum	Genbank (Table S7)
Vigna radiata	Genbank (Table S7)
Glycyrrhiza uralensis	Genbank (Table S7)
Vicia sativa	Genbank (Table S7)
Artemisia annua	Genbank (Table S7)
Arabidopsis thaliana	Genbank (Table S7)
Medicago truncatula	Genbank (Table S7)
Oryza sativa	Genbank (Table S7)
Physcomitrella patens	Genbank (Table S7)
Populus trichopoda	Genbank (Table S7)
Lotus japonicus	Genbank (Table S7)
Selaginella moellendorffii	Genbank (Table S7)

Solanum lycopersicum	Genbank (Table S7)
Solanum tuberosum	Genbank (Table S7)
Borya sphaerocephala	1KP
Chlorogalum pomeridianum	1KP
Curculigo sp.	1KP
Cyanastrum cordifolium	1KP
Eriospermum lancifolia	1KP
Freycinetia multiflora	1KP
Goodyera pubescens	1KP
Helonia bullata	1KP
Hesperaloe parviflora	1KP
Johnsonia pubescens	1KP
Maianthemum canadense	1KP
Nolina atopocarpa	1KP
Oncidium spacelatum	1KP
Platanthera clavellata	1KP
Ruscus sp.	1KP
Sansevieria trifasciata	1KP
Xeronema callistemon	1KP
Xerophyllum asphodeloides	1KP
Xerophyta villosa	1KP
Yucca brevifolia	1KP
Yucca filamentosa	1KP
Amborella trichopoda	Amobrella Genome Project
Musa acuminata	Cirad (D'Hont *et al.* 2012)
Nelumbo nucifera	GB: GCA_000365185.2
Phoenix dactylifera	JCGR (Al-Mssallem *et al.* 2013)
Acorus americana	MonAToL
Anemarrhena asphodeloides	MonAToL
Aphyllanthes monspeliensis	MonAToL

Asparagus asparagoides	MonAToL
Behnia reticulata	MonAToL
Cypripedium acaule	MonAToL
Doryanthes excelsa	MonAToL
Hemiphylacus alatostylus	MonAToL
Ixiolirion sp.	MonAToL
Lanaria larata	MonAToL
Lilium superbum	MonAToL
Paphiopedilum callosum	MonAToL
Phragmidpedium lindleyannum	MonAToL
Hosta venusta	MonAToL; GB: SRX116252
Chlorophytum rhizopendulum	MonAToL; GB: SRX116253
Aquilegia coerulea	Phytozome
Arabidopsis thaliana	Phytozome
Carica papya	Phytozome
Fragaria vesca	Phytozome
Glycine max	Phytozome
Medicago truncatula	Phytozome
Mimulus guttatus	Phytozome
Oryza sativa	Phytozome
Populus trichopoda	Phytozome
Selaginella moellendorffii	Phytozome
Solanum lycopersicum	Phytozome
Solanum tuberosum	Phytozome
Sorghum bicolor	Phytozome
Theobroma cacao	Phytozome
Vitis vinifera	Phytozome
Veratrum californicum	This Study

Table S9. LS/MS rooting medium was prepared with the following concentrations and brought to a final pH of 5.75.

Macronutrients	Supplier	Final Concentration (mg/l)
NH_4NO_3	Phytotechnology laboratories	1650
KNO_3	Sigma	1900
$MgSO_4$ x $7H_2O$	Sigma	370
KH_2PO_4	Sigma	170
$CaCl_2$ x $2H_2O$	Sigma	440
Iron		
Na_2EDTA x $2H_2O$	Sigma	37.3
$FeSO_4$ x $7H_2O$	Sigma	27.8
Micronutrients		
H_3BO_3	Sigma	6.2
$MnSO_4$ x H_2O	Phytotechnology laboratories	16.9
$ZnSO_4$ x $7H_2O$	Sigma	8.6
KI	Sigma	0.83
Na_2MoO_4 x $2H_2O$	Phytotechnology laboratories	0.25
$CuSO_4$ x $5H_2O$	Sigma	0.025
$CoCl_2$ x $6H_2O$	Sigma	0.025
Vitamines		
Thiamine HCl	Sigma	0.1
Nicotinic acid	Sigma	0.5
Pyroxidine HCl	Sigma	0.5
Myo-inositol	Sigma	100
Other		
Sucrose	Phytotechnology laboratories	30000
1-Naphthaleneacetic acid	Phytotechnology laboratories	0.5
Gelzan	Phytotechnology laboratories	3000

Table S10. LC-MS/MS Q-TRAP 4000 method parameters.

	Q1 Mass	Q3 Mass	Dwell (msec)	Declustering Potential (V)	Collision Energy (V)
Cyclopamine	412	321	100	120	40
	412	394	100	120	40

22-Keto-26-hydroxycholesterol	417	271	100	70	30
	417	253	100	70	30
22-Hydroxy-26-aminocholesterol	418	400	100	70	30
	418	382	100	70	30
Verazine	398	253	100	70	60
	398	159	100	70	60

Table S11. Primer sequences.

Number	Primer Position and Application	5' – 3' Sequence
1	5' UTR outer for 2646	CAAGTCGTGATTGATGGCTTTAGAAGGCA
2	5' UTR inner for 2646	TGGATCTCTGAAGCCATGAATCGCTAGTA
3	3' UTR outer for 2646	TCCTATAACCATTTATTTCTCGTAACC
4	3' UTR inner for 2646	ATGCAGAGAGCAATATTACAACCCAA
5	5' gene specific primer (GSP) used to amplify 2646 and introduce NotI RS	TGAGCGGCCGCATGGCGATGGAGCTCTTATTGTTG
6	3' used to amplify 2646 and introduce BamHI RS	TGAGGATCCTTAGTCTCCGAGGGGGCGAACTT
7	5' UTR for 13284	AGAAAGAAAGAGAGAGAGATGACTCCA
8	3' UTR for 13284	TAGAGAAAGACTGCTTGAATTTTTCAGGCAA
9	5' used to amplify 13284 and introduce PstI RS	GGGCTGCAGATGACTCCACTAGTTGTTCTCTTC
10	3' used to amplify 13284 and introduce XbaI RS	GGGTCTAGATTATTGGAGGAGCGAAAGCCG
11	5' used to amplify 12084 and introduce BglII RS	GGGAGATCTATGGGATCCACTGAGGCGCCTGTATC
12	3' used to amplify 12084 and introduce EcoRI RS	GGGGAATTCTTAAGTTGCGGTATTCTGAGACTGG

13	5’ outer RACE for 12709	GCACTGGAGCACGAGGACACTGA
14	5’ inner RACE for 12709	GGACACTGACATGGACTGAAGGAGTA
15	Outer gene specific internal reverse for 12709 5’ RACE	TGGTCCACCACGTGCGGCCGCGACGAGA
16	Inner Gene specific internal reverse for 12709 5’ RACE	GAACCAGGTGAGCGCCGAGGGGGTGGT
17	5’ used to amplify 12709 and introduce BglII RS	GGGAGATCTATGGATCTACCCTCCGCCTC
18	3’ used to amplify 12709 and introduce EcoRI RS	GGGGAATTCCTAACACCCTCTCTTCCTCTCCTTG
19	5’ used to amplify tomato GABAT2 and introduce PstI RS	CACACTGCAGATGGCCAAGACTAATGGATTTATG
20	3’ used to amplify tomato GABAT2 and introduce XbaI RS	CACATCTAGATTACTTCTTCTGAGACTTTAATTCTTCCA
21	5’ used to amplify 674 and introduce PstI RS	CACACTGCAGATGTTCTCAAGGCAAGCTACAG
22	3’ used to amplify 674 and introduce XbaI RS	CACATCTAGACTACTTCTGCTTCTGAGACTTGAGCT
23	5’ used to amplify 12084 for insertion into pET28a and introduce NdeI RS	TTTCATATGGGATCCACTGAGGCGCCTGTAT
24	3’ used to amplify 12084 for insertion into pET28a and introduce EcoRI RS	GGGGAATTCTTAAGTTGCGGTATTCTGAGACTGGAGCT

Table S12. PCR Parameters.

2646 for TOPO cloning, VC2646 for pVL1392 cloning, VC12709 for pVL1392 cloning, VC13284 TOPO cloning	98°C for 30 sec, then 35 cycles of 98°C for 10 sec, 60°C for 30 sec, 72°C for 1 min 30 sec, and a final 10 min extension at 72°C

12709 5’ RACE	98°C for 30 sec, then 35 cycles of 98°C for 10 sec, 65°C for 30 sec, 72°C for 1 min and a final 10 min extension at 72°C
13284 for pVL1392 cloning	98°C for 30 sec, then 35 cycles of 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min 25 sec, and a final 10 min extension at 72°C
12084 for pVL1392 cloning	98°C for 30 sec, then 35 cycles of 98°C for 10 sec, 60°C for 30 sec, 72°C for 1 min, and a final 10 min extension at 72°C
Tomato GABAT and 674 for pVL1392 cloning	98°C for 30 sec, then 35 cycles of 98°C for 10 sec, 62°C for 30 sec, 72°C for 45 sec, and a final 10 min extension at 72°C
12084 for pET28a cloning	98°C for 30 sec, then 35 cycles of 98°C for 10 sec, 63°C for 30 sec, 72°C for 2 min, and a final 10 min extension at 72°C

Table S13. Viral combinations for *in vivo* production of metabolites in *S. frugiperda* Sf9 cells.

Combination	**Viruses**
Combination 1	CYP90B27, CPR
Combination 2	CYP94N1, CPR
Combination 3	CYP90G1, CPR
Combination 4	CYP90B27, CYP94N1, CPR
Combination 5	CYP90B27, CYP90G1, CPR
Combination 6	CYP90B27, CYP94N1, CYP90G1, CPR
Combination 7	CYP90B27, CYP94N1, CYP90G1, GABAT1, CPR
Combination 8	CYP90B27, CYP94N1, GABAT1, CPR

Combination 9	CYP90B27, CYP94N1, CYP90G1, GABAT2, CPR
Combination 10	CYP90B27, CYP94N1, CYP90G1, *S. lycopersicum* GABAT2, CPR

2.9.3 Supporting data

Data S1. Data access for sequence information.

Illumina reads and assembled transcriptome retrieval of *Veratrum californicum, Narcissus* sp. *aff. pseudonarcissus*, and *Colchicum autumnale.*

Each species was assembled by the NCGR and the raw reads, assembled transcriptome, and readme files can be found at: http://www.medplantrnaseq.org/.

Data access for 1KP and MonATol Projects.

Data for the 1KP can be accessed at https://sites.google.com/a/ualberta.ca/onekp/

Data for the MonATol project is currently unpublished and access to the sequence information was courtesy of James H. Leebens-Mack (jleebensmack@plantbio.uga.edu). All sequences and alignments used in this analysis are deposited in Dryad: DOI: doi:10.5061/dryad.n9s7q.

Data S2. NMR designations for 22-keto-cholesterol and 22-keto-26-hydroxycholesterol in MeOD.

22-keto-cholesterol. ^{1}H NMR (600 MHz, CD_3OD): δ 5.34 (m, 1H, H-6), 3.40 (m, 1H, H-3), 2.52-2.62 (m, 2H, H-20, H-23), 2.45 (m, 1H, H-23), 2.23 (m, 1H, H-4), 1.95-2.07 (m, 2H, H-7, H-12), 1.88 (dt, *J* = 13.0, 3.5 Hz, 1H, H-1), 1.78 (m, 1H, H-2), 1.45-1.67 (m, 8H, H-2, H-7, H-11, H-15, H-16, H-17, H-25), 1.41 (m, 1H, H-24), 1.31 (m, 1H, H-12), 1.19 (m, 1H, H-16), 1.10 (d, *J* = 7.0 Hz, 3H, H-21), 1.04-1.09 (m, 2H, H-1, H-14), 1.03 (s, 3H, H-19), 0.98 (td, *J* = 11.0, 5.0 Hz, 1H, H-9), 0.90 (m, 6H, H-26, H-27), 0.76 (s, 3H, H-18); ^{13}C NMR (150 MHz, CD_3OD): δ 217.8 (C-22), 141.9 (C-5), 122.1 (C-6), 72.1 (C-3), 57.3 (C-14), 53.6 (C-17), 51.4 (C-9), 50.2 (C-20), 43.4 (C-13), 42.6 (C-4), 40.8 (C-12), 40.4 (C-23), 38.3 (C-1), 37.7 (C-10), 36.8 (C-8), 33.1 (C-24), 32.7 (C-7), 31.8 (C-2), 28.6 (C-25), 28.2 (C-16), 25.8 (C-15), 22.5 (C-27), 22.0 (C-11), 19.6 (C-19), 16.7 (C-21), 13.2 (C-26), 12.2 (C-18).

22-keto-26-hydroxycholesterol. ^{1}H NMR (600 MHz, CD_3OD): δ 5.34 (br d, *J* = 5.1 Hz, 1H, H-6), 3.30-3.42 (m, 3H, H-3, H-26), 2.58-2.66 (m, 2H, H-20, H-23), 2.46 (ddd, *J* = 17.6, 9.2, 5.9 Hz, 1H, H-23), 2.22 (m, 2H, H-4), 1.94-2.03 (m, 2H, H-7, H-12), 1.88 (dt, *J* = 13.6, 13.3 Hz, 1H, H-1), 1.79 (m, 1H, H-2), 1.44-1.70 (m, 10H, H-2, H-7, H-8, H-11, H-15, H-16, H-17, H-24, H-25), 1.30-1.34 (m, 2H, H-12, H-24), 1.21 (m, 1H, H-16), 1.15 (m, 1H, H-15), 1.11 (d, *J* = 7.0 Hz, 3H, H-21), 1.04-1.10 (m, 2H, H-1, H-14), 1.03 (s, 3H, H-19), 0.98 (td, *J* = 11.2, 5.0 Hz, 1H, H-9), 0.91 (d, *J* = 6.6 Hz, 3H, H-27), 0.76 (s, 3H, H-18); ^{13}C NMR (150 MHz, CD_3OD): δ 217.0 (C-22), 141.2 (C-5), 122.1 (C-6), 72.1 (C-3), 67.7 (C-27), 57.1 (C-14), 53.3 (C-17), 51.3 (C-9), 50.2 (C-20), 43.2 (C-13), 42.7 (C-4), 40.6 (C-12), 40.1 (C-23), 38.3 (C-1), 37.4 (C-10), 35.9 (C-25), 32.7 (C-7), 32.5 (C-8), 31.9 (C-2), 28.1 (C-16), 27.4 (C-24), 25.4 (C-15), 21.8 (C-11), 19.6 (C-19), 16.7 (C-21), 16.6 (C-27), 12.0 (C-18).

2.9.4 Supporting methods

Method S1. Plant material and RNA extraction.

Veratrum californicum plant material was obtained from wild populations in northern Utah. Tissue culture was initiated from wild collected seed and grown in the dark at 24°C on a combination of Linsmaier and Skoog vitamins (Linsmaier E.M. 1965) and Murashige and Skoog media(Murashige 1962) supplemented with 0.5 mg/l 1-naphthaleneacetic acid (Sigma). Refer to Table S9 for detailed media components. RNA extraction for each tissue (bulb, flower, leaf, fall rhizome, spring rhizome, fall root, green shoot, white shoot, and tissue culture samples) was performed as previously described (Johnson *et al.* 2012) (protocol 13). RNA quantity and integrity were evaluated with a NanoDrop 2000 (Thermo Scientific) and a Bioanalyzer 2100 (Agilent Technologies) prior to cDNA library preparation.

Method S2. *Veratrum californicum* metabolite extraction for quantitation by LC-MS/MS.

Extracts were prepared by grinding frozen plant tissue in liquid nitrogen followed by 5 minutes of vortexing in 70% ethanol added in a 200 µl to 100 mg w/v ratio. Samples were subject to centrifugation for 10 minutes (14,000 X g) at room temperature and the supernatant filtered through a 0.2 µm PTFE membrane (Millipore) prior to injection. Extracts were diluted 10-10,000 fold with 70% ethanol, depending on alkaloid concentration, prior to LC-MS/MS analysis (refer to LC-MS/MS protocol in Experimental Procedures).

Method S3. Transcriptome assembly and determination of relative contig expression.

cDNA library construction, Illumina paired-end sequencing, and *de novo* transcriptome assembly for *Veratrum californicum* were performed at the National Center for Genome

Resources (Santa Fe, New Mexico). For the transcriptome assembly, 50 bp paired-end Illumina reads for each tissue were first examined for gross abnormalities and poor sequence quality and trimmed with the FASTX Toolkit. (**http://hannonlab.cshl.edu/fastx_toolkit**). Quality control was as previously described (Kilgore *et al.* 2014).The reads were 5′ and 3′ quality trimmed using a Phred score of 15 to eliminate noisy reads. Subsequently, short contig assembly was performed using the de Bruijn graph-based assembler ABySS (Assembly by Short Sequences) several times with varying k-mer lengths to generate 20 sets of synthetic ESTs (Expressed Sequence Tags) with lengths between 100 – 500 base pairs (Birol *et al.* 2009, Simpson *et al.* 2009). ABySS scaffolder was used to scaffold the synthetic ESTs and GapCloser from SOAPdenovo (Short Oligonucleotide Analysis Package) to close the NNN gap spacers (Li *et al.* 2010). Lastly, the assembly was completed by combining the obtained scaffolds using MIRA (Mimicking Intelligent Read Assembly) in the EST assembly mode (Chevreux 2005)**.** Post processing included translational predictions for each contig using ESTSCAN (Iseli *et al.* 1999, Lottaz *et al.* 2003), protein product motif annotation written to GFF3 files (General Feature Format), and determination of expression data by alignment analysis of the trimmed reads to the assembled contigs using BWA (Burrows-Wheeler Aligner) (Li and Durbin 2009). To further enable comparison of gene expression between various tissues, the number of reads aligned to each contig was normalized by dividing by the total number of reads from the respective tissue sample. Functional annotations to each predicted protein sequence were obtained using Pfam (Punta *et al.* 2012), Superfamily (Gough *et al.* 2001), and Uniprot (UniProt 2013).

Method S4. Transcriptome dataset interrogation using Haystack and PlantTribes.
Haystack (http://haystack.mocklerlab.org/) input parameters included a value of 20 for fall root, fall rhizome, spring rhizome and bulb and a value of 1 for leaf, flower, white shoot, green shoot, and tissue culture 1- and 2 weeks after transfer to new media. A correlation cutoff value of 0.7 was used instead of the default of 0.8 to avoid missing true positives.

Gene family circumscriptions in PlantTribes were calculated using the similarity-based clustering of gene models from *Arabidopsis thaliana* (v. 7), *Carica papaya* (v.1), *Populus trichocarpa* (v. 1), *Medicago truncatula* (v. 1), *Oryza sativa* (v. 5), *Sorghum bicolor* (v. 1), *Selaginella moellendorffii* (v. 1), *Physcomitrella patens* (v. 1), and *Chlamydomonas reinhardtii* (v. 1) genome annotations using TribeMCL (Enright *et al.* 2003, Enright *et al.* 2002). Translated transcript assemblies for *Veratrum californicum, Narcissus* sp*. aff. pseudonarcissus* (daffodil), and *Colchicum autumnale* (autumn crocus) were sorted into the resulting gene family clusters using BLAST followed by multiple sequence alignment and gene tree estimation. In addition to this MCL (Markov CLuster

algorithm) clustering approach, we developed a complete minimal representative dataset from all available plant species of cytochrome P450 genes relevant to alkaloid biosynthesis to identify candidate cytochromes P450.

Multiple sequence alignment and phylogenetic tree estimation were done on these relevant tribes and gene families using the MAFFT (Multiple Alignment using Fast Fourier Transform) alignment software (Katoh *et al.* 2009) and RAxML (Randomized Axelerated Maximum Likelihood) for maximum likelihood tree generation (Stamatakis 2006).

The RNA-seq transcriptome assembly sequences from *C. autumnale* and *Narcissus* were included in the tribe clustering steps (Supporting Data S1). These two species are also lillioid monocots but do not produce cyclopamine (but instead make the unrelated alkaloids colchicine and galanthamine, respectively), *C. autumnale* and *Narcissus* sequences were included in order to facilitate identification of tribe clusters that only contain *Veratrum californicum* genes, which were then assigned a higher priority for biochemical validation.

Selection criteria were established to score and sort the resulting clades, positive criteria were given a +1 each while penalizations were scored as a -1 each. A given clade was scored on the percentage of clade members that significantly co-localized with cyclopamine (e.g. present in the Haystack output dataset). Clades that did not contain any genes that fit the Haystack model were penalized. Clades containing genes that were *not* significantly co-localized with the alkaloid were penalized. Lastly, clades that contain genes from species that *do not* produce cyclopamine incurred a score penalty and were not chosen for initial biochemical characterization. These criteria were combined to score and rank the clades that contain Haystack output gene members to identify the clade(s) with the highest likelihood of containing genes that function in the steroid alkaloid biosynthesis pathway.

Method S5. Construction of viral expression vectors.

Candidate contigs obtained from Haystack analysis were subjected to BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and global alignments to homologous, experimentally characterized gene sequences with the CLC Main Workbench 6.8, for prediction of the open reading frame. Where the reading frame appeared incomplete, Rapid Amplification of cDNA Ends (RACE) was used to obtain the complete coding sequence. *Veratrum californicum* cDNA was prepared from root RNA extracts using M-MLV Reverse Transcriptase (Invitrogen)

according to manufacturer's instructions. All primer sequences and PCR programs can be found in Tables S11 and S12, respectively.

The cDNAs encoding CYP90B27 (accession numbers KJ869252, KJ869253), CYP90G1 (accession numbers KJ869258-KJ869261), GABAT1 (accession numbers KJ869262-KJ869264) and γ-aminobutyrate transaminase 2 (GABAT2) (accession number KJ869265) were determined to be full length. CYP90B27 and CYP90G1 were amplified by Polymerase Chain Reaction (PCR) from cDNA with Phusion DNA polymerase (New England Biolabs) using primers 1-4 and 7-8, respectively, and initially ligated into the pCR-Blunt II-TOPO vector (Invitrogen). Two rounds of amplification were required for CYP90B27 by nested PCR. Subsequently, CYP90B27 was amplified from pCR-Blunt II-TOPO with primers 5 and 6 introducing NotI/BamHI restriction sites into the PCR products at the 5' and 3' ends of the open reading frame. The amplified product and pVL1392 Baculovirus transfer vector (BD Biosciences) were digested with NotI/BamHI and ligated together using Rapid Ligase (Promega). Ligated constructs were transformed into *E. coli* DH5α competent cells. CYP90G1 was amplified with primers 9 and 10, introducing PstI/XbaI restriction sites at the 5' and 3' end of the open reading frame. The amplified product, along with pVL1392, was digested with PstI/XbaI and subject to ligation and transformation.

GABAT1 and GABAT2 were directly amplified from cDNA using primers 11, 12 incorporating BglII/EcoRI restriction sites at the 5' and 3' end of the open reading frame and 21, 22, incorporating PstI/XbaI restriction sites at the 5' and 3' end of the open reading frame, respectively. GABAT1 and pVL1392 were subject to restriction digest with BglII/EcoRI preceding ligation and transformation. GABAT2 was digested with XbaI/PstI, preceding ligation and transformation.

RACE was required to determine the 5' sequence of CYP94N1 gene (accession numbers KJ869254-KJ869257). RACE ready cDNA was prepared using the GeneRacer Kit (Invitrogen) according to manufacturer's instructions using *V. californicum* root RNA. Primers 13 and 15 were used for PCR (round 1), followed by amplification using primers 14 and 16 (round 2). Resulting RACE fragments were cloned into PCR-Blunt II-TOPO. The full-length gene was directly amplified from *V. californicum* root cDNA with primers 17 and 18, incorporating BglII/EcoRI restriction sites at the 5' and 3' end of the open reading frame. The amplified product was digested with Bglll/EcoRI and ligated into pVL1392 digested with the same

enzymes. Each characterized *V. californicum* contig and subsequent enzyme designation can be found in Table S6.

The cDNA encoding GABA transaminase isozyme 2 from *Solanum lycopersicum* (tomato GABAT2) implicated in steroid alkaloid biosynthesis (accession number AY240230) was isolated from *S. lycopersicum* using the Qiagen RNA-easy kit for RNA extraction followed by cDNA synthesis as described above. Tomato GABAT2 was amplified by PCR using Primers 19 and 20, incorporating PstI/XbaI sites at the 5' and 3' end of the open reading frame. The amplified product and pVL1392 were subject to restriction digest with PstI/XbaI and ligated together, preceding transformation.

Method S6. Virus co-transfection, amplification, and protein production.
Each pVL1392 expression construct was independently co-transfected with the Baculogold Linearized Baculovirus (BD Biosciences) into *S. frugiperda* Sf9 cells according to manufacturer's instructions. Sf9 cells were maintained as previously described (Gesell *et al.* 2011). Virus amplification and protein production proceeded as previously described (Gesell *et al.* 2009). Each cytochrome P450 virus construct was co-expressed with virus containing *E. californica* cytochrome P450 reductase (CPR) while the GABAT1 was produced by single infection. Sf9 cell cultures were also infected with several constructs in parallel. Combinations of each virus used in multiple infections can be found in Table S13. Equal volumes for each virus were used in the multiple infections and adjusted to a total viral volume of 2.5 ml.

Method S7. Extraction of multiple infections for Sf9 *in vivo* product production.
Baculovirus infections were carried out for production of each enzymatic product in *S. frugiperda* Sf9 cells and collected as stated in Method S6. 1 ml each of Sf9 cells expressing the various combinations of virus were extracted with 2 volumes of ethyl acetate by vortexing (1 min), centrifugation (16,000 x g; 2 min), and were taken to dryness under N_2. Samples were either derivatized with 40 µl of Sylon HTP and injected onto the GC-MS using the protocol stated in Experimental Procedures or were re-suspended in 50 µl of 80% methanol and analyzed by LC/MS-MS according to the protocol in Experimental Procedures.

Method S8. Assays to clarify order of enzymatic transformations.
Assay for GC-MS: Cytochrome P450 enzyme assay conditions were identical to those stated above using *S. frugiperda* Sf9 cell suspensions with the following modifications. First, 12 assays each containing CYP90G1 + CPR, CYP94N1 + CPR, or control cytochrome P450 + CPR and each with pure 22(*R*)-hydroxycholesterol were allowed to incubate overnight at 30°C. Like

assays were pooled, extracted 3 times with 2 volumes ethyl acetate, dried under N_2, and re-suspended in 180 µl of 25% DMSO. Extracts containing the enzymatic product of the 22-hydroxycholesterol 26-hydroxylase/ oxidase + CPR and 22(*R*)-hydroxycholesterol were divided equally and used as substrate for 6 assays containing CYP90G1 + CPR and 6 assays containing control cytochrome P450 + CPR. Extracts containing the enzymatic product of CYP90G1 + CPR and 22(*R*)-hydroxycholesterol were divided and used as substrate in 6 assays containing 22-hydroxycholesterol 26-hydroxylase/ oxidase + CPR and 6 assays containing control cytochrome P450 + CPR. Control P450 + CPR assay was run in parallel, treated identically and added to another control P450 assay. Assays were allowed to incubate for 20 min at 30°C then stopped by addition of 20 µl of 20% TCA with vortexing. Like assays were pooled, extracted, derivatized, and analyzed by GC-MS using the protocol stated in Experimental Procedures. Refer to Figure S11 a for an overview of the experiment.

Assay for LC-MS/MS: All assays used crude Sf9 cell suspensions. Enzyme assays started with a combination of CYP90B27 + CPR and CYP94N1 + CPR (8 individual reactions) in parallel to CYP90B27 + CPR and CYP90G1 + CPR (8 reactions). Assays were extracted, and added to CYP94N1 + CPR, CYP90G1 + CPR, or GABAT1 for several possible enzyme combinations (4 reactions each). Like samples were pooled, extracted, and added to 2 reactions each with enzyme not yet utilized previously. Refer to Figure S11 b for an overview of the experiment. Samples were taken at each step post extraction for LC-MS/MS analysis and run by the protocol stated in Experimental Procedures.

Method S9. Enzymatic product purification for NMR and high resolution MS for structure elucidation.

Large-scale 750 ml *S. frugiperda* Sf9 cultures were grown expressing viral combinations 5-7 (Table S13) of the *Veratrum californicum* enzymes as previously described (Gesell *et al.* 2009). Cells were collected after three days and re-suspended in 10 ml of 100 mM tricine pH 7.4/ 5 mM thioglycolic acid; then extracted 3 times with 2 volumes of hexane or ethyl acetate. The remaining aqueous supernatant was extracted once with 1 volume of hexane or ethyl acetate. Extracts for each infection were then pooled, dried under N_2, and re-suspended in 5 ml of absolute methanol.

The extracts were purified on a Waters HPLC system equipped with a 2707 autosampler, 1525 binary pump, 2998 photodiode array detector, and Waters Fraction Collector III. In some cases,

samples were cleaned up by Solid Phase Extraction (SPE), before HPLC purification. For HPLC, extracts were concentrated to 500 µl and then injected in 50 µl portions onto a Phenomenex Gemini C-18 NX column (150 X 2.00 mm, 5 µm) with the same solvents used for LC-MS/MS as described in Experimental Procedures with the following binary gradient: Solvent B was held at 20% for 2 min, then 2-11 min 20-30% B, 11-18 min 30-100% B, 18-30 min 100% B, 30-31 min 100-20% B, and held at 20% B for an additional 5 minutes. The flow rate was 0.5 ml/min; 0.5 ml fractions were collected. The resulting fractions were then analyzed by GC-MS or LC-MS/MS as described in Experimental Procedures, and selected samples were analyzed by NMR or by high resolution MS. NMR spectra were acquired in MeOD at 600 MHz on a BrukerAvance 600 MHz spectrometer equipped with a BrukerBioSpin TCI 1.7 mm MicroCryoProbe. Proton, gCOSY, ROESY, gHSQC, and gHMBC spectra were acquired; ^{13}C chemical shifts were obtained from the HSQC and HMBC spectra. Chemical shifts are reported with respect to the residual non-deuterated MeOD signal. Refer to Data S2 for NMR designations for 22-keto-cholesterol and 22-keto-26-hydroxycholesterol. For high resolution MS, samples were diluted 1:10 in 80% acetonitrile:water (LC-MS grade) containing 0.1% formic acid and infused into an LTQ-Orbitrap Velos Pro (Thermo-Fisher Scientific, San Jose, CA) using a Triversa Nanomate (Advion, Ithaca, NY). Data were collected in positive ion mode, detected in the Orbitrap at a nominal resolution setting of 60,000 at m/z 400. Precursors were determined with a wide SIM scan (m/z 385-430). Precursors were isolated in the ion-trap and transferred to the HCD cell for fragmentation at 35 NCE (m/z 418) and 50 NCE (m/z 398). Data were analyzed manually using the Qualbrowser application of Xcalibur (Thermo-Fisher Scientific, San Jose, CA).

Method S10. Phylogenetic analysis of cytochrome P450 enzymes across species using deep transcriptome sequence data from 1KP and MonAToL projects.

Sequences for P450 genes were identified and pulled from various sources. Transcriptome data for monocotyledon taxa were taken from the MonAToL project (DEB-0830020) and the 1KP project (https://sites.google.com/a/ualberta.ca/onekp/home). Taxa were selected to represent major clades but were not exhaustive of the sampling available. We also included genome sequences from the 22 sequenced land plant genomes used by the *Amborella* genome project to represent eudicots and other angiosperms (Amborella Genome 2013). The final taxa used in reconstructing the P450 gene tree can be found in Table S8. Additionally, a P450 sequence from *S. cerevisiae* was used as the outgroup (GenBank Accession: U34636.1).

Transcriptomic data was assembled using Trinity r2013-02-25 and resulting assemblies were filtered using FPKM (fragments per kilobase of exon per million fragments mapped). Sequences where an isoform represented less than 1% of all reads mapping to a gene were discarded. Assemblies were translated using the ORF estimation Transdecoder r20131110 software packaged with Trinity. Blastp (Camacho *et al.* 2009) was used to blast genomic and transcriptomic amino acid sequences using the *S. lycopersicum* GAME4, GAME6, GAME7, and GAME8 (Refer to Table S7 for accession numbers) and *Veratrum californicum* CYP90B27, CYP90G1, and CYP94N1 (Refer to Table S7 for accession numbers) P450 amino acid sequence. Best blast hits were identified using an initial e-value threshold of 1e-10 followed by a minimum overlapping length of 85% between the query and subject sequences. These filtering criteria resulted in putative P450 genes of at least 1017 basepairs in length for all taxa.

Amino acid sequences for all putative P450 sequences were aligned using MAFFT v.7.029b (Katoh and Standley 2014) under default settings. Trees were estimated using RAxML v.8.0.22 (Stamatakis 2006) under the GTR + Γ evolutionary model with 500 bootstrap replicates and *S. cerevisiae* as the outgroup.

Chapter 3: Production of *Veratrum californicum* secondary metabolites in *Camelina sativa* seed

3.1 Summary

Economically feasible systems suitable for heterologous production of secondary metabolites originating from complex species are in demand. Established systems such as *Escherichia coli* and *Saccharomyces cerevisiae* are not always suitable for expression of plant and animals genes due to several differences including mechanisms of post-translational modifications and cellular organization. An emerging oilseed crop, *Camelina sativa*, has recently been engineered to produce novel oil profiles, jet fuel precursors, and small molecules of industrial interest. One medicinally relevant secondary metabolite, cyclopamine, is currently used to produce medicines in clinical trials, and the future supply of this potential cancer treatment is uncertain. The sole source of cyclopamine, *Veratrum californicum*, is slow growing and not amendable to cultivation. In order to establish *C. sativa* as a system for the production of medicinally relevant compounds, we introduced four genes from *V. californicum* involved in steroid alkaloid biosynthesis. These four genes produce verazine, the hypothesized precursor to cyclopamine. Herein, we successfully engineered *C. sativa* to produce verazine, as well as other *V. californicum* secondary metabolites, in seed.

3.2 Significance statement

Verazine was produced in *Camelina sativa* seeds by heterologous expression of *Veratrum californicum* genes CYP90B27, CYP94N1, CYP90G1, GABAT1 and the *Arabidopsis thaliana* gene GAD2. This is the first report of verazine production in a heterologous system. In addition, this is the first report of *C. sativa* use for the production of medicinally relevant compounds.

3.3 Author contributions

I, (Megan M. Augustin) in addition to Ashutosh K. Shukla and Cynthia Holland constructed vectors for transformation. I performed all plant transformations and seed collection. I, with the help of Linna Han, extracted seeds for metabolite analysis. I ran all samples on the LC-MS/MS and analyzed the results. I generated all data for all figures. I produced all figures and wrote the manuscript.

3.4 Introduction

Industrial production of plant derived natural products for medicinal use can be unachievable by either traditional chemical synthesis or agricultural techniques. Chemical synthesis techniques

often require harsh conditions and result in toxic byproducts. The complex stereochemistry of many natural products can also complicate synthesis. In addition, direct extraction of medicinal compounds from source plant species is not always a viable alternative, as many produce low natural yields and/or not amendable to cultivation (Atanasov *et al.* 2015). The development of a low input heterologous eukaryotic system is in demand for high value medicinal compounds. One potential production system is the emerging agricultural oilseed crop *Camelina sativa*.

C. sativa is an attractive alternative to current heterologous production systems for plant natural products such as *Escherichia coli* and *Saccharomyces cerevisiae* because of its molecular compatibility with other plant species, as well as its low input requirements. The close phylogenetic relatedness to the well-studied model organism *Arabidopsis thaliana*, in addition to ease of transformation and a sequenced genome, offers a foundation of genetic and molecular tools that can be utilized for metabolic manipulation (Kagale *et al.* 2014, Lu and Kang 2008). Complication arising from codon usage and post-translational modifications would be minimal in a plant production system as opposed to the distantly related microbial alternatives. Advantages to cultivation of *C. sativa* include low water- and nutrient input, as well as its compatibility with current agricultural practices (Putnam *et al.* 1993). In addition, its short growing season (ca. 100 days) along with cold tolerance has shown its potential for growth and harvest prior to the normal growing season, therefore maximizing land use without effecting normal crop production (Putnam *et al.* 1993).

C. sativa is under extensive investigation for its use in biofuel, industrial oil/lubricants, and high value metabolite production. Genetic manipulation of fatty acid content by heterologous expression of novel *Cuphea* FatB genes has altered fatty acid content in *C. sativa* by increasing medium chain fatty acids, therefore mimicking the hydrocarbons present in Jet A fuel (Kim *et al.* 2015). Transgenic *C. sativa* was also shown to accumulate valuable liquid wax esters by addition of wax ester synthase from *Mus musculus* or *Simmondsia chinensis* and fatty acyl-CoA reductase from *Mus musculus* or *Marinobacter aquaeolei* (Iven *et al.* 2015). RNAi suppression *of C. sativa* type 1 diacylglycerol acyltransferase led to accumulation of 3-acetyl-1,2-diacyl-sn-glycerols, triacylglycerols with attractive chemical properties including reduced freezing point and viscosity (Liu *et al.* 2015). *C. sativa* has also been under investigation for the use in aquafeeds as a replacement for vegetable oils that lack omega-3 long chain polyunsaturated fatty acids (ω3 LC-PUFA). Accumulation of the highly desired ω3 LC-PUFA eicosapentaenoic acid and docosahexaenoic acid has also been achieved by heterologous expression of yeast, algae, and additional species' genes involved in the delta6-desaturase pathway (Petrie *et al.*

2014, Ruiz-Lopez *et al.* 2014). These high yielding ω3 LC-PUFA lines of *C. sativa* have been shown to successfully replace fish oil in aquaculture (Betancor *et al.* 2015a, Betancor *et al.* 2015b). Non-genetically modified *C. sativa* has also shown promise in aquaculture (Hixson *et al.* 2013, Hixson *et al.* 2014a, Hixson *et al.* 2014b).

In addition to advances in oil composition, genetically manipulated *C. sativa* has shown promise for the accumulation and storage of other highly valuable compounds. The production of the mono- and sesquiterpenes limonene and cadinene (respectively), in transgenic *C. sativa* could provide a robust and renewable source for these industrially useful terpenes, in addition to their potential use in jet fuel (Augustin *et al.* 2015a). *C. sativa* was also engineered to produce the biodegradable polymer poly-3-hydroxybutyrate (PHB). PHB can be used as a renewable component in bio-plastics, among other industrial uses (Malik *et al.* 2015). Moreover, production of these metabolites in seeds lends to easy harvest and long term storage, desirable traits for large scale production. We hypothesize, based on the previous success of metabolic engineering of *C. sativa*, that we can use *C. sativa* as a production system to produce medicinally useful small metabolites in seed.

The stereochemically complex steroidal alkaloid cyclopamine is one medicinal compound currently under clinical investigation whose supply is predicted to fall short of potential demands upon FDA approval (Heretsch *et al.* 2010). Cyclopamine is best known for its teratogenic effects in lambs born to pregnant ewes that ingested the source plant *Veratrum californicum* (Keeler and Binns 1968). Cyclopamine's mode of action is inhibition of the hedgehog pathway, mainly active during embryonic development, by direct binding to the transmembrane receptor Smoothend (Chen *et al.* 2002). Cyclopamine, and its semi-synthetic analog IPI-926, has shown promise in cancer therapy where mutations in the hedgehog pathway cause over activation leading to tumor growth, including pancreatic cancer, medulloblastoma, basal cell carcinoma, leukemia, colon cancer, and small cell lung cancer (Bahra *et al.* 2012, Batsaikhan *et al.* 2014, Berman *et al.* 2002, Jimeno *et al.* 2013, Lin *et al.* 2010, Olive *et al.* 2009, Taipale *et al.* 2000, Tremblay *et al.* 2009, Watkins *et al.* 2003). Cultivation of *V. californicum* for production of cyclopamine has yet to be successful, and cell cultures grow slowly and fail to produce substantial amount of cyclopamine (Song *et al.* 2014). Four genes have been discovered in the hypothesized biosynthetic pathway to cyclopamine (Augustin *et al.* 2015b). Together, the enzymatic products of these genes convert cholesterol to the predicted cyclopamine steroidal alkaloid precursor verazine. By transforming these four genes in addition to Arabidopsis glutamate decarboxylase 2 (GAD2), an enzyme required to produce the co-substrate γ-

aminobutyric acid (GABA) in seed, we have successfully engineered *C. sativa* to accumulate verazine in seed (Turano and Fang 1998).

3.5 Results

3.5.1 Vector construction, plant transformation, and confirmation of construct integration

Vector construction for plant transformation consisted of two parts. The first gene was cloned into an initial vector that contained a seed specific promoter and terminator. Next, the expression cassette (consisting of the promoter, gene, and terminator) was cut out by restriction enzyme digest and ligated into the multiple cloning site (MCS) of the plant expression vector pRSe3. Multiple expression cassettes were then cloned into the MSC of one plant transformation vector which allowed several genes to be transformed at once. Two plant expression vectors were constructed. One contained three genes involved in *V. californicum* steroid alkaloid biosynthesis and was designated CTOG (**C= C**holesterol 22-hydroxylase (CYP90B27), **T=** 22-Hydroxycholesterol-26-al **T**ransaminase (GABAT1), **O=** 22-Hydroxycholesterol 26-hydroxylase/**O**xidase (CYP94N1, and **G= G**lutamate decarboxylase 2 (GAD2)). The second construct contained four genes involved in *V. californicum* steroid alkaloid biosynthesis and was designated CXTOG (**X =** 22-Hydroxy-26-aminocholesterol 22-**oX**idase (CYP90G1)). GAD2 was included in both vector constructs in order to increase the concentration of GABA in seed, a co-substrate required by GABAT1. The vector pRSe3 is a modified Ti (tumor inducing) plasmid from *Agrobacterium* that allows for genomic integration of the DNA contained between the left and

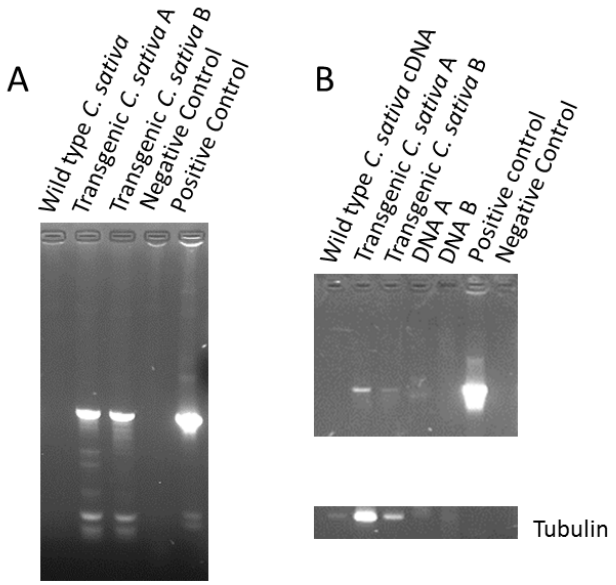

Figure 1. Confirmation of construct integration and expression. A) PCR products for GAD2 gene using DNA extracted from selected *C. sativa* lines transformed with CYP90B27, CYP94N1, CYP90G1, GABAT1 from *V. californicum* and GAD2 from *A. thaliana* were run on a 1% agarose gel. A negative control assay lacking template DNA and a positive control assay with purified vector for DNA template were included. B) PCR products for GAD2 synthesized using cDNA generated from *C. sativa* RNA were analyzed on a 1% agarose gel. DNA from two individual wild type *C. sativa* plants were run as control in addition to a positive control containing purified vector DNA as template and a negative control without DNA.

right border sequences (referred to as T-DNA), but does not allow for bacterial infection. The vector also contains a Kanamycin resistance gene for selection in bacteria and the DsRed gene for selection in *C. sativa* seed. *C. sativa* plants were transformed as previously described (Lu and Kang 2008). Initial confirmation of T-DNA integration into transformed *C. sativa* was achieved by visualization of DsRed by illuminating seeds with a green LED light and observing fluorescence through a red filter. After transgenic seeds were grown to the next generation, PCR analysis on selected CXTOG transgenic *C. sativa* confirmed integration of the construct (Figure 1 A). Gene expression was also confirmed by PCR analysis on cDNA derived from total developing seed RNA (Figure 1 B). In addition, production of metabolites as described below supported integration and expression of the transgenes.

3.5.2 4000 QTRAP analysis for detection of V. californicum *metabolites in transgenic C. sativa seeds*

Initial metabolite screening was performed on wild type and transgenic CTOG and CXTOG *C. sativa* seed extracts using LC-MS/MS with a 4000 QTRAP for detection of verazine and previously identified metabolites produced by the transgenes (Augustin *et al.* 2015b). Accumulation of 22-hydroxy-26-aminocholesterol, in addition to three unknown compounds with the same mass and similar fragmentation pattern, was detected in transgenic CTOG plants containing CYP90B27, CYP94N1, GABAT1, and GAD2, but lacking CYP90G1 (Figure 2). 22-Keto-26-hydroxycholesterol and verazine were not detected. In CXTOG plants containing CYP90B27, CYP94N1, CYP90G1, GABAT1, and GAD2, 22-hydroxy-26-aminocholesterol was not detected, but 22-keto-26-hydroxycholesterol was found to accumulate. In addition to this compound, another peak eluted 1 minute later and appeared to have the same mass and similar fragmentation pattern (Figure 3). A peak for verazine was detected by a more sensitive scan (MRM scan) and appeared slightly above background and was not detected in the wild type plants, but further analysis and verification was desired (Figure 4). In addition, a novel peak eluted after verazine with the same fragment ion and was also not detected in the wild type or Sf9 extracts.

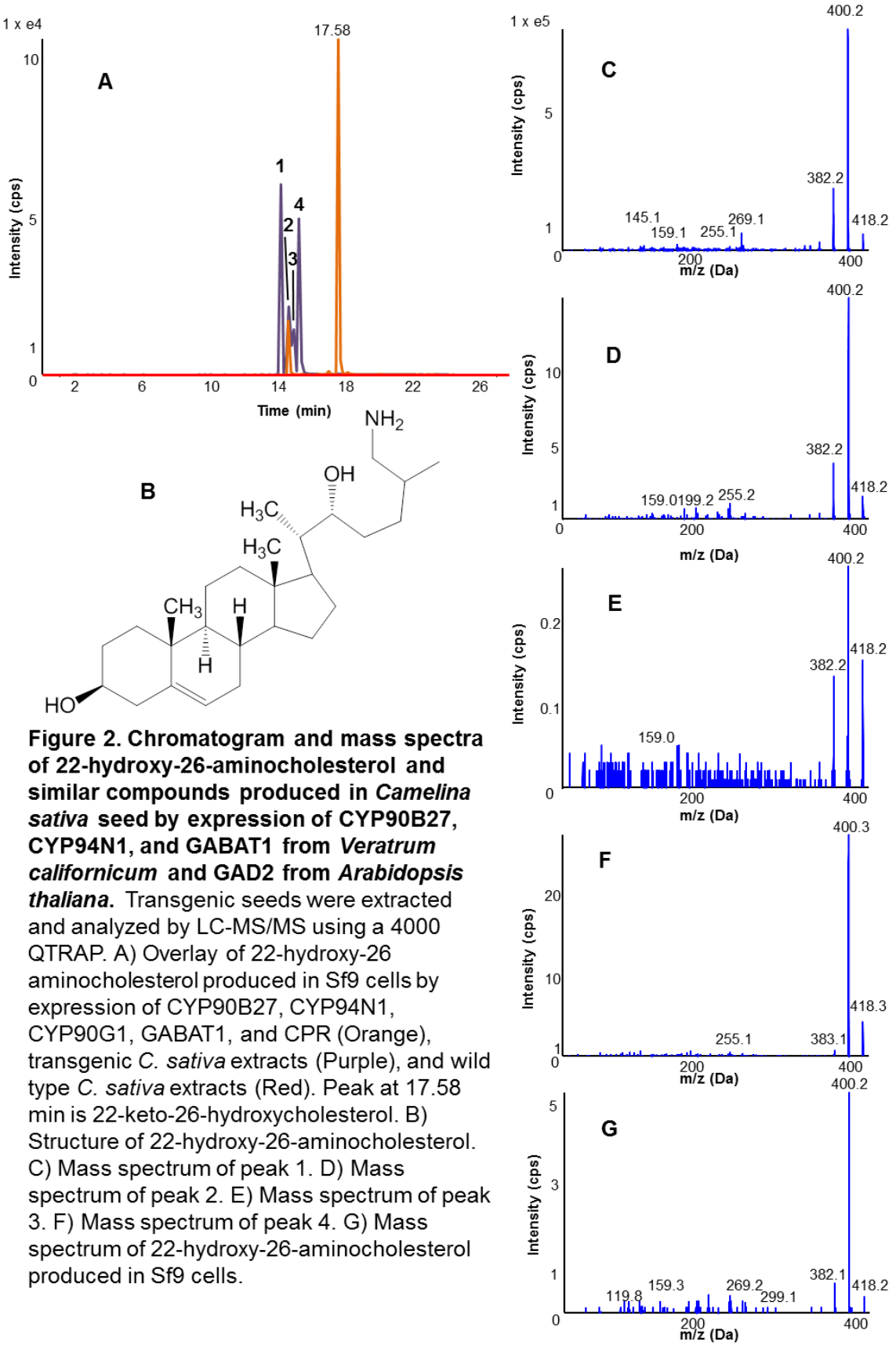

Figure 2. Chromatogram and mass spectra of 22-hydroxy-26-aminocholesterol and similar compounds produced in *Camelina sativa* seed by expression of CYP90B27, CYP94N1, and GABAT1 from *Veratrum californicum* and GAD2 from *Arabidopsis thaliana*. Transgenic seeds were extracted and analyzed by LC-MS/MS using a 4000 QTRAP. A) Overlay of 22-hydroxy-26 aminocholesterol produced in Sf9 cells by expression of CYP90B27, CYP94N1, CYP90G1, GABAT1, and CPR (Orange), transgenic *C. sativa* extracts (Purple), and wild type *C. sativa* extracts (Red). Peak at 17.58 min is 22-keto-26-hydroxycholesterol. B) Structure of 22-hydroxy-26-aminocholesterol. C) Mass spectrum of peak 1. D) Mass spectrum of peak 2. E) Mass spectrum of peak 3. F) Mass spectrum of peak 4. G) Mass spectrum of 22-hydroxy-26-aminocholesterol produced in Sf9 cells.

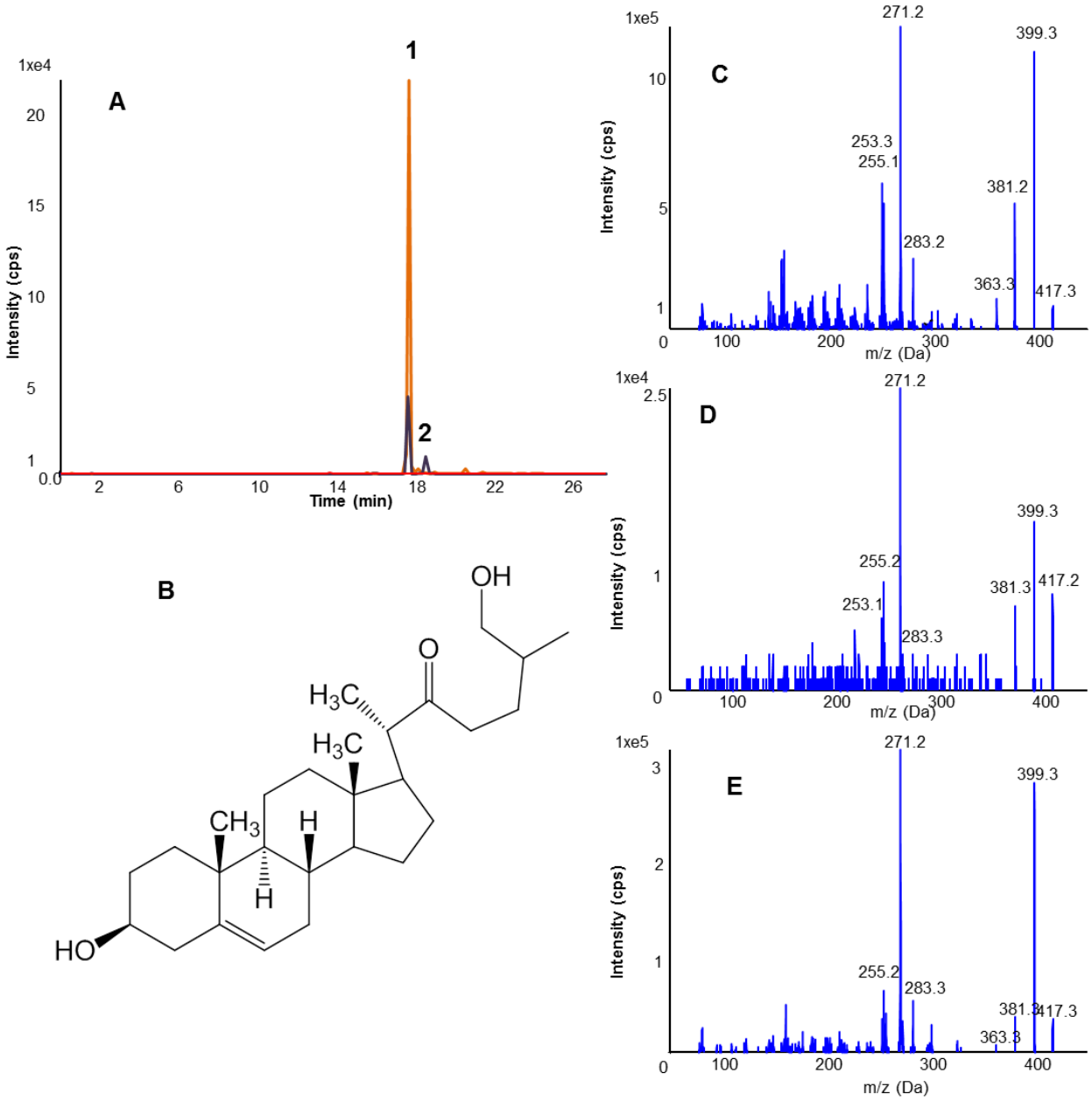

Figure 3. Targeted LC-MS/MS analysis of 22-keto-26-hydroxycholesterol from transgenic *Camelina sativa* expressing CYP90B27, CYP94N1, CYP90G1, and GABAT1 from *Veratrum californicum* and GAD2 from *Arabidopsis thaliana*. Extracts were analyzed by 4000 QTRAP using EPI scan for 417 *m/z*. A) Overlay of 22-keto-26-hydroxycholesterol extracted from heterologous expression of CYP90B27, CYP94N1, CYP90G1, CPR, and GABAT1 in Sf9 insect cells (Orange), extracts of *C. sativa* expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and GAD2 (Purple), and wild type *C. sativa* (Red). B) Structure of 22-keto-26-hydroxycholesterol. C) Mass spectrum of 22-keto-26-hydroxycholesterol from Sf9 cells (peak 1). D) Mass spectrum of 22-keto-26-hydroxycholesterol from transgenic *C. sativa* (peak 1). E) Mass spectrum of peak 2 in transgenic *C. sativa.* Selected fragment ions are shown.

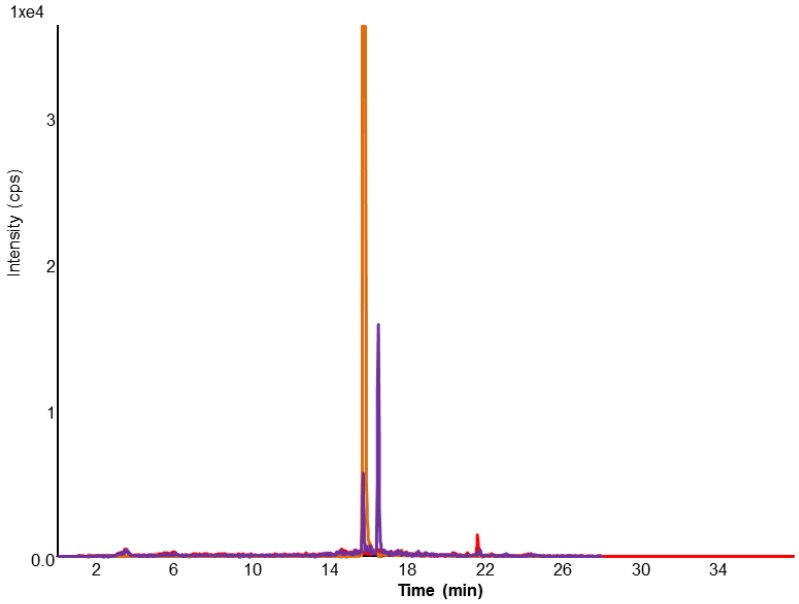

Figure 4. Targeted LC-MS/MS analysis of verazine from transgenic *Camelina sativa* expressing CYP90B27, CYP94N1, CYP90G1, and GABAT1 from *Veratrum californicum* and GAD2 from *Arabidopsis thaliana*. Extracts of Sf9 cells expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and CPR (Orange), transgenic *C. sativa* expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and GAD2 (Purple), and wild type *C. sativa* (Red) were analyzed by LC-MS/MS using a 4000 QTRAP with MRM scan for ion 398.300/159.200 for increased sensitivity.

3.5.3 Q-Exactive analysis for verazine detection

In order to verify the accumulation of verazine in CXTOG transgenic *C. sativa* seeds, samples were analyzed by high-resolution mass spectrometry. The increased sensitivity of the Q-Exactive enhanced detection of verazine and allowed us to obtain accurate mass in addition to a mass spectrum (Figure 5). The detection of a peak with mass 398.3418 in the transgenic *C. sativa* CXTOG extracts coincided with verazine produced in Sf9 cells. Both had only 0.25 ppm and 0.5 ppm mass measurement error, respectively, when compared to the calculated accurate mass of 398.3417. In addition, the defining fragment peak 126.1279 was found in the transgenic plant samples as well as the verazine from Sf9 cells. This mass corresponds to only 1.5 ppm mass measurement error when compared to the calculated accurate mass of 126.1277. All mass measurement errors fall far below the 5 ppm limit for determining compound identity. In addition to verazine, another metabolite with the same mass and similar fragmentation pattern was also detected in the transgenic *C. sativa* extracts (Figure 5 F).

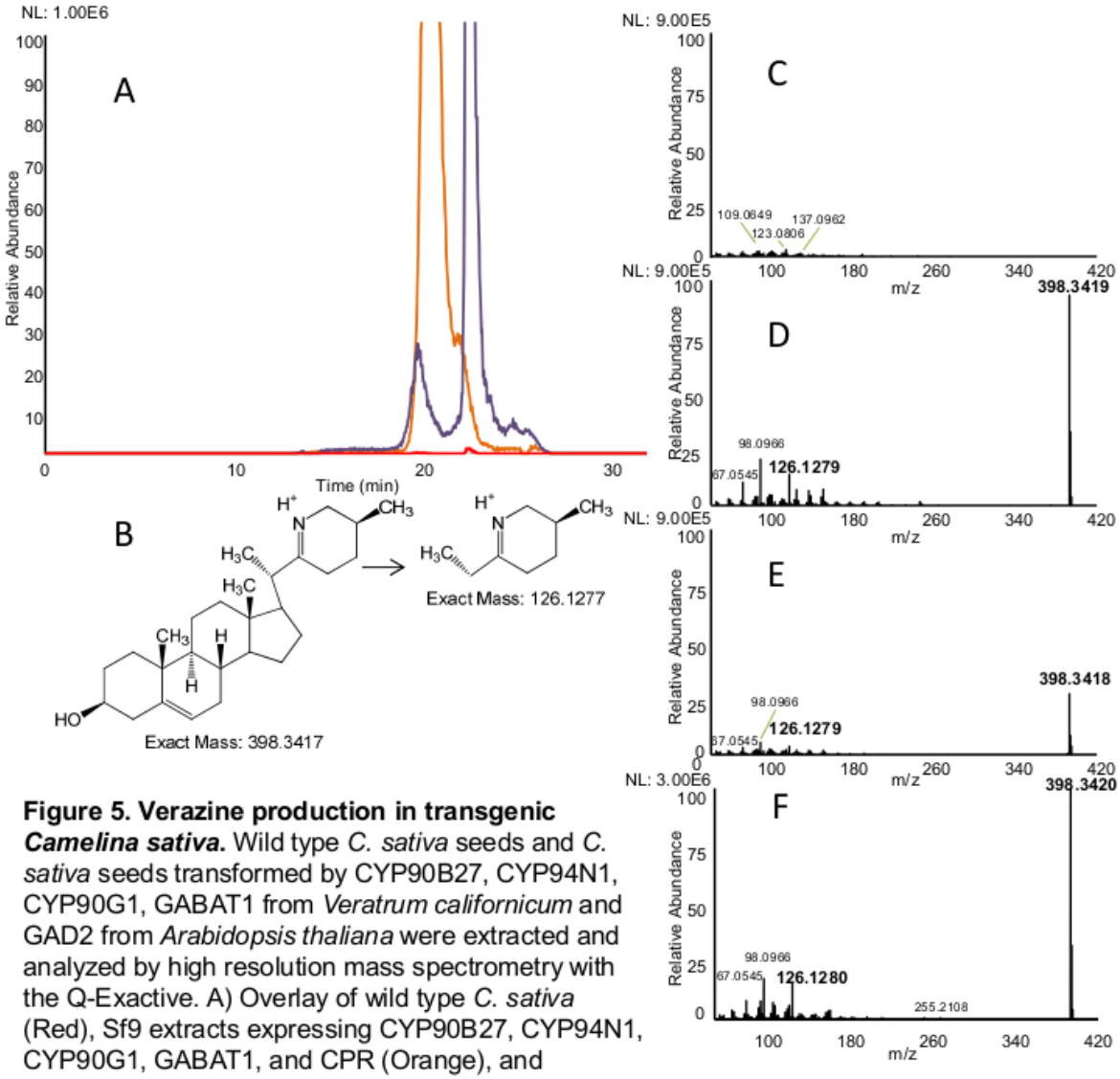

Figure 5. Verazine production in transgenic *Camelina sativa*. Wild type *C. sativa* seeds and *C. sativa* seeds transformed by CYP90B27, CYP94N1, CYP90G1, GABAT1 from *Veratrum californicum* and GAD2 from *Arabidopsis thaliana* were extracted and analyzed by high resolution mass spectrometry with the Q-Exactive. A) Overlay of wild type *C. sativa* (Red), Sf9 extracts expressing CYP90B27, CYP94N1, CYP90G1, GABAT1, and CPR (Orange), and transgenic *C. sativa* expressing CYP90B27, CYP94N1, CYP90G1, GABAT1 and GAD2 (Purple). B) Calculated exact mass of verazine and key fragment ion. MS2 Fragmentation pattern for time point 19.9 min, elution time for verazine, are shown on the right including C) wild type D) Sf9 extracts, and E) transgenic *C. sativa.* F) Mass spectrum for peak eluting at 22 min in transgenic *C. sativa*. Spectra were filtered for exact mass of 398.3417 with a mass tolerance of 20 ppm. Exact mass of verazine and key fragment are indicated in bold.

3.6 Discussion

C. sativa engineered to express CYP90B27, CYP94N1, CYP90G1, GABAT1, and GAD2 was shown to accumulate verazine, the hypothesized intermediate to the antineoplastic cyclopamine. In addition to verazine, 22-keto-26-hydroxycholesterol was also detected. Additional peaks with the same masses and similar fragmentation patterns occurred at different retention times *in planta* for both of these compounds, suggesting the possibility of alternative stereochemical isomers produced in *C. sativa*. Reduction/elimination of these side products may be achieved by introducing sequential cyclopamine biosynthetic enzymes once they are discovered. As described for *A. thaliana* engineered with three genes to produce dhurrin, a cyanogenic glucoside from Sorghum, many unintended side products accumulated in transgenic plants containing only two genes. When the third gene was introduced, and the pathway thereby complete, the unintended side products were no longer detected and the end product was greatly enhanced (Kristensen *et al.* 2005).

Despite the absence of true verazine quantitation (due to a lack of available standard), it was observed that the production was quite low. Several possibilities exist for further engineering of *C. sativa* for increased verazine and sequential metabolite production in the future. As mentioned earlier, additional pathway enzymes may increase yield and drive the production of metabolites with the correct stereochemistry. Another possibility would be to increase the production of cholesterol in *C. sativa,* a precursor molecule for steroid alkaloids. Cholesterol has been found as only a minor steroid in *C. sativa,* so this may be limiting for metabolite production (Mansour *et al.* 2014). Overexpression of oxidosqualene cyclases such as cycloartenol synthase or lanosterol synthase may increase the available cholesterol (Ohyama *et al.* 2009). In addition, upregulation of bottleneck enzymes in the mevalonic acid pathway may enhance yield. As previously demonstrated by Augustin *et al*.2015aq, overexpression of the rate limiting enzyme 1-deoxy-d-xylulose-5-phosphate synthase (DXS) in *C. sativa* engineered to produce limonene and cadinene significantly increased terpene production. A deeper understanding of cholesterol biosynthesis in plants will enhance decisions in this regard. Engineering of this pathway into another system such as *S. cerevisiae* is also a viable option, but would not include the benefits of a *C. sativa* production system as described earlier.

Cultivation of *C. sativa* as an agricultural crop is industrially and environmentally attractive due to its low input requirements and chemical properties. Efficient use of cultivatable land by growing *C. sativa* prior to the typical growing season will provide an opportunity for metabolite and fuel production without jeopardizing food crops or compromising undeveloped wilderness

(Bansal and Durrett 2015). In addition, after metabolite extraction, the leftover material can be utilized as biofuels or other industrial applications, wasting little material (Asomaning *et al.* 2014). Herein, we demonstrated the production of *V. californicum* secondary metabolites in *C. sativa,* highlighting the possibilities for future engineering and industrial production.

3.7 Experimental procedures

Cloning and plant transformation

CYP90B27, CYP94N1, CYP90G1, and γ-aminobutyrate transaminase 1 (GABAT1) were cloned from *Veratrum californicum* total root cDNA and glutamate decarboxylase 2 (GAD2) was cloned from *Arabidopsis thaliana* (L.) Heyn., ecotype Columbia (Col-0) total leaf cDNA. *V. californicum* plant tissues were obtained from wild collection in Northern Utah, USA and the Arabidopsis tissue was obtained onsite (Donald Danforth Plant Science Center Greenhouse, St. Louis, MO, USA). Total RNA from *V. californicum* was extracted as previously described (Johnson *et al.* 2012). Arabidopsis RNA was extracted using an RNeasy Plant Mini Kit (Qiagen). cDNA synthesis followed using MMLV-RT. Genes were first ligated into cassette vectors to provide each with a seed specific promoter and terminator before transfer into the plant expression vector pRSe3, a binary vector with kanamycin and DsRed selection markers. pRSe3 is based on pRSe2, with an enhanced multiple cloning site (Augustin *et al.* 2015a). The following genes were ligated into each cassette vector providing the following seed specific promoters/terminators: CYP90B27 (Oleosin/Oleosin); CYP94N1 (Napin/Glycinin); GABAT1 (Napin/Glycinin); CYP90G1 (Napin/Glycinin); and GAD2 (Glycinin/Glycinin). Final constructs were made both with (CXTOG) and without (CTOG) CYP90G1. Plants were transformed as previously described (Lu and Kang 2008). Transgenic seeds were screened as previously described (Augustin *et al.* 2015a). Verazine, 22-keto-26-hydroxycholesterol, and 22-hydroxy-26-aminocholesterol were heterologously produced in Sf9 insect cells as previously described (Augustin *et al.* 2015b).

DNA extraction, RNA extraction, and PCR

DNA was extracted from leaf tissue of selected plants grown from seeds exhibiting red fluorescence for verification of construct integration using the DNeasy Plant Mini Kit (Qiagen). PCR was performed with primers GAD2_F_NdeI (CACACATATGGTTTTGACAAAAACCGCAACGA), and GAD2_RC_NotI (CACAGCGGCCGCTTAGCACACACCATTCATCTTCTT) and the following temperature program parameters: 2 min 95 °C, 1 cycle; 30 s 95 °C, 30 s 52 °C, 1 min 72 °C, 30 cycles; final

extension at 72 °C for 5 min. RNA was extracted from developing seeds with the RNeasy Plant Mini Kit (Qiagen). cDNA synthesis followed using the RNA to cDNA EcoDry kit (Clontech) and expression was confirmed by amplification of the GAD2 gene using the same primers and PCR cycle described above. Amplification of tubulin was included as a control.

Seed extraction

T2 (CTOG) and T3 (CXTOG) generation seeds were extracted based upon the protocol in the acyl-lipid metabolism chapter in The Arabidopsis Book (Li-Beisson *et al.* 2013). Hot isopropanol (1.5 ml, 75°C) was added to 15-20 mg of seeds and incubated for 15 min. Seeds were then crushed with a glass rod followed by the addition of chloroform and H_2O (0.75 ml and 0.3 ml, respectively). Next, samples were sonicated in a sonication bath for 10 min followed by robust shaking for 1 hour at room temperature (RT). Tubes were then vortexed for 10 seconds and centrifuged (1,500 x g, 2 min, RT). The liquid was moved to a fresh tube and the remaining tissue was re-extracted once with 2 ml of chloroform:methanol (2:1) and once with 1 ml of chloroform:methanol (2:1). Supernatants were combined prior to addition of 0.5 ml 1 M KCl. Samples were vortexed, centrifuged, and the aqueous upper phase was removed. This extraction was repeated once by the addition of 1 ml H_2O. The organic phase was dried to completion under N_2. Extracts were re-suspended in 80% methanol and filtered with a 0.2 µ PTFE filter (Millipore) before LC-MS/MS analysis.

LC-MS/MS analysis

Seed extracts were first analyzed by LC-MS/MS using a 4000 QTRAP (ABSciex) as previously described (Augustin *et al.* 2015b). Selected seed extracts were additionally analyzed by high-resolution mass spectrometry using a Q-Exactive (Thermo Fisher) coupled to an Agilent1200 microLC system. Samples were separated by a PLRP-S column (100 x 0.5 mm, 100Å, 3µ; Higgins Analytical) with a flow rate of 15 µl/min and the following solvent/gradient system: solvent A (0.05% formic acid/0.01% ammonium hydroxide v/v in H_2O; solvent B (0.05% formic acid/0.01% ammonium hydroxide v/v in 90% acetonitrile]: Solvent B was held at 10% for 4 min, then 4-10 min 10-40% B, 10-18 min 40-45% B, 18-20 min 45-100% B, 20-21 min 100% B, 21-22 100-10% B and held at 10% B for an additional 10 minutes. Results were analyzed using Analyst 1.5 (Applied Biosystems) for data generated with the 4000 QTRAP and Xcalibur 3.0.36 (Thermo Scientific) for data generated by Q-Exactive.

Chapter 4: Concluding remarks and future perspectives

With the conclusion of this work, four genes have been discovered in the biosynthesis of steroid alkaloids in *V. californicum*. The enzymes encoded by these genes synthesize verazine from cholesterol, a hypothesized intermediate of the antineoplastic cyclopamine. Verazine itself could serve as a precursor molecule for the semi-synthesis of a variety of steroid alkaloids. The enzymatic introduction of nitrogen can reduce the number of steps required to produce other potentially useful steroid alkaloids that would otherwise require harsh chemical reactions (Atanasov *et al.* 2015). Also of note, verazine has been shown to be a significant anti-fungal agent, which could be useful for the development of novel anti-fungal treatments (Kusano *et al.* 1987).

The biosynthetic pathway to verazine, at least in this heterologous system, was clarified. The pathway found herein is slightly different than what was predicted previously. The *V. californicum* metabolite 22-keto-26-hydroxycholesterol does not appear to play a role in the formation of verazine. It was also clarified that the C-22 position was hydroxylated first, followed by hydroxylation at position C-26. We also established the nitrogen source for GABAT1 as GABA, not arginine. In addition, oxidation at position C-22 must occur after nitrogen incorporation, thus driving the formation of the piperidine ring. We also found that gene expression does not entirely overlap with the accumulation of cyclopamine. These enzymes may be involved in the formation of additional steroid alkaloids in *V. californicum*, as the plant makes several types which may all rely on these initial biosynthetic genes. These discoveries, however, are only the beginning for future cyclopamine production.

Continued research is necessary for complete elucidation of this pathway. As the semi-synthetic analog IPI-926 proceeds in clinical trials, and the prospects of cyclopamine and its derivatives remain high for cancer treatment, the demand for this compound will outweigh the supply. Several more complex steps are required to form cyclopamine from verazine including hydroxylation at position C-16 followed by closure of the E-ring. Hydroxylation at position C-12 must follow, preceding the notable ring rearrangement for formation of the C-*nor*-D-*homo* skeleton. Opening and closing of the E-ring must occur with the addition of an oxygen atom before cyclopamine can be formed as well. Once these genes have been discovered, future work on cyclopamine production at an industrial level can be continued. The synthesis of IPI-926, however, could possibly initiate from one of the cyclopamine precursors, and full pathway elucidation may not be required.

In addition, it has been demonstrated that the *V. californicum* genes CYP90B27, CYP94N1, CYP90G1, and GABAT1 are functional in the heterologous expression system *C. sativa*. Accumulation of verazine was confirmed in transgenic seed. Continued efforts are necessary to increase production and overcome bottlenecks, but the basis for production has been established. Moreover, the use of *C. sativa* as a heterologous production system for medicinally useful metabolites has been explored and is supported by the work herein. Future work on *C. sativa*, in conjunction with the vast knowledge base from the close relative *A. thaliana*, can help achieve greater accumulation of metabolites by manipulation of primary metabolic pathways. New technologies, such as RNAi, can help downregulate specific genes without eliminating them completely, allowing for the fine tuning of metabolic pathways for optimal metabolite production. *C. sativa* will, however, remain in the infancy stage as a production system unless more focus is placed on this crop and its potential is fully explored.

References

Adam, G., Schreiber, K., Tomko, J. and Vassova, A. (1967) [Verazine, a new Veratrum alkaloid with 22, 26-imino-cholestan structure]. *Tetrahedron*, **23**, 167-171.

Al-Mssallem, I.S., Hu, S., Zhang, X., Lin, Q., Liu, W., Tan, J., Yu, X., Liu, J., Pan, L., Zhang, T., Yin, Y., Xin, C., Wu, H., Zhang, G., Ba Abdullah, M.M., Huang, D., Fang, Y., Alnakhli, Y.O., Jia, S., Yin, A., Alhuzimi, E.M., Alsaihati, B.A., Al-Owayyed, S.A., Zhao, D., Zhang, S., Al-Otaibi, N.A., Sun, G., Majrashi, M.A., Li, F., Tala, Wang, J., Yun, Q., Alnassar, N.A., Wang, L., Yang, M., Al-Jelaify, R.F., Liu, K., Gao, S., Chen, K., Alkhaldi, S.R., Liu, G., Zhang, M., Guo, H. and Yu, J. (2013) Genome sequence of the date palm Phoenix dactylifera L. *Nature communications*, **4**, 2274.

Amborella Genome, P. (2013) The Amborella genome and the evolution of flowering plants. *Science*, **342**, 1241089.

Asomaning, J., Mussone, P. and Bressler, D.C. (2014) Two-stage thermal conversion of inedible lipid feedstocks to renewable chemicals and fuels. *Bioresource technology*, **158**, 55-62.

Atanasov, A.G., Waltenberger, B., Pferschy-Wenzig, E.M., Linder, T., Wawrosch, C., Uhrin, P., Temml, V., Wang, L., Schwaiger, S., Heiss, E.H., Rollinger, J.M., Schuster, D., Breuss, J.M., Bochkov, V., Mihovilovic, M.D., Kopp, B., Bauer, R., Dirsch, V.M. and Stuppner, H. (2015) Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnology advances*.

Augustin, J.M., Higashi, Y., Feng, X. and Kutchan, T.M. (2015a) Production of mono- and sesquiterpenes in Camelina sativa oilseed. *Planta*, **242**, 693-708.

Augustin, M.M., Ruzicka, D.R., Shukla, A.K., Augustin, J.M., Starks, C.M., O'Neil-Johnson, M., McKain, M.R., Evans, B.S., Barrett, M.D., Smithson, A., Wong, G.K., Deyholos, M.K., Edger, P.P., Pires, J.C., Leebens-Mack, J.H., Mann, D.A. and Kutchan, T.M. (2015b) Elucidating steroid alkaloid biosynthesis in Veratrum californicum: production of verazine in Sf9 cells. *The Plant journal : for cell and molecular biology*.

Bahra, M., Kamphues, C., Boas-Knoop, S., Lippert, S., Esendik, U., Schuller, U., Hartmann, W., Waha, A., Neuhaus, P., Heppner, F., Pietsch, T. and Koch, A. (2012) Combination of hedgehog signaling blockage and chemotherapy leads to tumor reduction in pancreatic adenocarcinomas. *Pancreas*, **41**, 222-229.

Bansal, S. and Durrett, T.P. (2015) Camelina sativa: An ideal platform for the metabolic engineering and field production of industrial lipids. *Biochimie*.

Batsaikhan, B.E., Yoshikawa, K., Kurita, N., Iwata, T., Takasu, C., Kashihara, H. and Shimada, M. (2014) Cyclopamine decreased the expression of Sonic Hedgehog and its downstream genes in colon cancer stem cells. *Anticancer research*, **34**, 6339-6344.

Behnsawy, H.M., Shigemura, K., Meligy, F.Y., Yamamichi, F., Yamashita, M., Haung, W.C., Li, X., Miyake, H., Tanaka, K., Kawabata, M., Shirakawa, T. and Fujisawa, M. (2013) Possible role of sonic hedgehog and epithelial-mesenchymal transition in renal cell cancer progression. *Korean journal of urology*, **54**, 547-554.

Berman, D.M., Karhadkar, S.S., Hallahan, A.R., Pritchard, J.I., Eberhart, C.G., Watkins, D.N., Chen, J.K., Cooper, M.K., Taipale, J., Olson, J.M. and Beachy, P.A. (2002) Medulloblastoma growth inhibition by hedgehog pathway blockade. *Science*, **297**, 1559-1561.

Berman, D.M., Karhadkar, S.S., Maitra, A., Montes De Oca, R., Gerstenblith, M.R., Briggs, K., Parker, A.R., Shimada, Y., Eshleman, J.R., Watkins, D.N. and Beachy, P.A. (2003) Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature*, **425**, 846-851.

Betancor, M.B., Sprague, M., Sayanova, O., Usher, S., Campbell, P.J., Napier, J.A., Caballero, M.J. and Tocher, D.R. (2015a) Evaluation of a high-EPA oil from transgenic in feeds for Atlantic salmon (L.): Effects on tissue fatty acid composition, histology and gene expression. *Aquaculture*, **444**, 1-12.

Betancor, M.B., Sprague, M., Usher, S., Sayanova, O., Campbell, P.J., Napier, J.A. and Tocher, D.R. (2015b) A nutritionally-enhanced oil from transgenic Camelina sativa effectively replaces fish oil as a source of eicosapentaenoic acid for fish. *Sci Rep*, **5**, 8104.

Binns, W., James, L.F., Shupe, J.L. and Everett, G. (1963) A Congenital Cyclopian-Type Malformation in Lambs Induced by Maternal Ingestion of a Range Plant, Veratrum Californicum. *American journal of veterinary research*, **24**, 1164-1175.

Binns, W., James, L.F., Shupe, J.L. and Thacker, E.J. (1962) Cyclopian-type malformation in lambs. *Arch Environ Health*, **5**, 106-108.

Binns, W., Keeler, R.F. and Balls, L.D. (1972) Congenital deformities in lambs, calves, and goats resulting from maternal ingestion of Veratrum californicum: hare lip, cleft palate, ataxia, and hypoplasia of metacarpal and metatarsal bones. *Clin Toxicol*, **5**, 245-261.

Binns, W., Shupe, J.L., Keeler, R.F. and James, L.F. (1965) Chronologic evaluation of teratogenicity in sheep fed Veratrum californicum. *Journal of the American Veterinary Medical Association*, **147**, 839-842.

Binns, W., Thacker, E.J., James, L.F. and Huffman, W.T. (1959) A congenital cyclopiantype malformation in lambs. *Journal of the American Veterinary Medical Association*, **134**, 180-183.

Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., Horsman, D.E., Connors, J.M., Gascoyne, R.D., Marra, M.A. and Jones, S.J. (2009) De novo transcriptome assembly with ABySS. *Bioinformatics*, **25**, 2872-2877.

Bryden, M.M. and Keeler, R.F. (1973) Proceedings: Effects of alkaloids of Veratrum californicum on developing embryos. *J Anat*, **116**, 464.

Bryden, M.M., Perry, C. and Keeler, R.F. (1973) Effects of alkaloids of Veratrum californicum on chick embryos. *Teratology*, **8**, 19-25.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC bioinformatics*, **10**, 421.

Chandler, C.M. and McDougal, O.M. (2014) Medicinal history of North American *Veratrum*. *Phytochemistry reviews : proceedings of the Phytochemical Society of Europe*, **13**, 671-694.

Chen, J.K., Taipale, J., Cooper, M.K. and Beachy, P.A. (2002) Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes & development*, **16**, 2743-2748.

Chevreux, B. (2005) MIRA: an automated genome and EST assembler (Heidelberg, R.-K.-U. ed.

Cochrane, C.R., Szczepny, A., Watkins, D.N. and Cain, J.E. (2015) Hedgehog Signaling in the Maintenance of Cancer Stem Cells. *Cancers (Basel)*, **7**, 1554-1585.

Cooper, M.K., Porter, J.A., Young, K.E. and Beachy, P.A. (1998) Teratogen-mediated inhibition of target tissue response to Shh signaling. *Science*, **280**, 1603-1607.

Corbit, K.C., Aanstad, P., Singla, V., Norman, A.R., Stainier, D.Y. and Reiter, J.F. (2005) Vertebrate Smoothened functions at the primary cilium. *Nature*, **437**, 1018-1021.

Corcoran, R.B. and Scott, M.P. (2001) A mouse model for medulloblastoma and basal cell nevus syndrome. *J Neurooncol*, **53**, 307-318.

Cordell, G.A. (1998) The alkaloids. Chemistry and biology. San Diego, Amsterdam: Academic Press, Elsevier.

D'Hont, A., Denoeud, F., Aury, J.M., Baurens, F.C., Carreel, F., Garsmeur, O., Noel, B., Bocs, S., Droc, G., Rouard, M., Da Silva, C., Jabbari, K., Cardi, C., Poulain, J., Souquet, M., Labadie, K., Jourda, C., Lengelle, J., Rodier-Goud, M., Alberti, A., Bernard, M., Correa, M., Ayyampalayam, S., McKain, M.R., Leebens-Mack, J., Burgess, D., Freeling, M., Mbeguie, A.M.D., Chabannes, M., Wicker, T., Panaud, O., Barbosa, J., Hribova, E., Heslop-Harrison, P., Habas, R., Rivallan, R., Francois, P., Poiron, C., Kilian, A., Burthia, D., Jenny, C., Bakry, F., Brown, S., Guignon, V., Kema, G., Dita, M., Waalwijk, C., Joseph, S., Dievart, A., Jaillon, O., Leclercq, J., Argout, X., Lyons, E., Almeida, A., Jeridi, M., Dolezel, J., Roux, N., Risterucci, A.M., Weissenbach, J., Ruiz, M., Glaszmann, J.C., Quetier, F., Yahiaoui, N. and Wincker, P. (2012) The banana (Musa acuminata) genome and the evolution of monocotyledonous plants. *Nature*, **488**, 213-217.

Dhakal, R., Bajpai, V.K. and Baek, K.H. (2012) Production of gaba (gamma - Aminobutyric acid) by microorganisms: a review. *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]*, **43**, 1230-1241.

Diaz Chavez, M.L., Rolf, M., Gesell, A. and Kutchan, T.M. (2011) Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy Argemone mexicana. *Arch Biochem Biophys*, **507**, 186-193.

Dierks, C., Grbic, J., Zirlik, K., Beigi, R., Englund, N.P., Guo, G.R., Veelken, H., Engelhardt, M., Mertelsmann, R., Kelleher, J.F., Schultz, P. and Warmuth, M. (2007) Essential role of stromally induced hedgehog signaling in B-cell malignancies. *Nature medicine*, **13**, 944-951.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792-1797.

Enright, A.J., Kunin, V. and Ouzounis, C.A. (2003) Protein families and TRIBES in genome sequence space. *Nucleic acids research*, **31**, 4632-4638.

Enright, A.J., Van Dongen, S. and Ouzounis, C.A. (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic acids research*, **30**, 1575-1584.

Epstein, E.H. (2008) Basal cell carcinomas: attack of the hedgehog. *Nature reviews. Cancer*, **8**, 743-754.

Gailani, M.R., Stahle-Backdahl, M., Leffell, D.J., Glynn, M., Zaphiropoulos, P.G., Pressman, C., Unden, A.B., Dean, M., Brash, D.E., Bale, A.E. and Toftgard, R. (1996) The role of the human homologue of Drosophila patched in sporadic basal cell carcinomas. *Nature genetics*, **14**, 78-81.

Gan, K.H., Lin, C.N. and Won, S.J. (1993) Cytotoxic principles and their derivatives of Formosan Solanum plants. *Journal of natural products*, **56**, 15-21.

Gesell, A., Chavez, M.L., Kramell, R., Piotrowski, M., Macheroux, P. and Kutchan, T.M. (2011) Heterologous expression of two FAD-dependent oxidases with (S)-tetrahydroprotoberberine oxidase activity from Arge mone mexicana and Berberis wilsoniae in insect cells. *Planta*, **233**, 1185-1197.

Gesell, A., Rolf, M., Ziegler, J., Diaz Chavez, M.L., Huang, F.C. and Kutchan, T.M. (2009) CYP719B1 is salutaridine synthase, the C-C phenol-coupling enzyme of morphine biosynthesis in opium poppy. *The Journal of biological chemistry*, **284**, 24432-24442.

Gorlin, R.J. (1995) Nevoid basal cell carcinoma syndrome. *Dermatol Clin*, **13**, 113-125.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of molecular biology*, **313**, 903-919.

Grothe, T., Lenz, R. and Kutchan, T.M. (2001) Molecular characterization of the salutaridinol 7-O-acetyltransferase involved in morphine biosynthesis in opium poppy Papaver somniferum. *The Journal of biological chemistry*, **276**, 30717-30723.

Hagel, J.M. and Facchini, P.J. (2010) Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nature chemical biology*, **6**, 273-275.

Hahn, H., Wicking, C., Zaphiropoulous, P.G., Gailani, M.R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A.B., Gillies, S., Negus, K., Smyth, I., Pressman, C., Leffell, D.J., Gerrard, B., Goldstein, A.M., Dean, M., Toftgard, R., Chenevix-Trench, G., Wainwright, B. and Bale, A.E. (1996) Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell*, **85**, 841-851.

Heretsch, P., Tzagkaroulaki, L. and Giannis, A. (2010) Cyclopamine and hedgehog signaling: chemistry, biology, medical perspectives. *Angewandte Chemie*, **49**, 3418-3427.

Hixson, S.M., Parrish, C.C. and Anderson, D.M. (2013) Effect of replacement of fish oil with camelina (Camelina sativa) oil on growth, lipid class and fatty acid composition of farmed juvenile Atlantic cod (Gadus morhua). *Fish Physiol Biochem*, **39**, 1441-1456.

Hixson, S.M., Parrish, C.C. and Anderson, D.M. (2014a) Changes in tissue lipid and fatty acid composition of farmed rainbow trout in response to dietary camelina oil as a replacement of fish oil. *Lipids*, **49**, 97-111.

Hixson, S.M., Parrish, C.C. and Anderson, D.M. (2014b) Full substitution of fish oil with camelina (Camelina sativa) oil, with partial substitution of fish meal with camelina meal, in diets for farmed Atlantic salmon (Salmo salar) and its effect on tissue lipids and sensory quality. *Food chemistry*, **157**, 51-61.

Huang, F.C. and Kutchan, T.M. (2000) Distribution of morphinan and benzo[c]phenanthridine alkaloid gene transcript accumulation in Papaver somniferum. *Phytochemistry*, **53**, 555-564.

Hui, C.C. and Angers, S. (2011) Gli proteins in development and disease. *Annu Rev Cell Dev Biol*, **27**, 513-537.

Incardona, J.P., Gaffield, W., Kapur, R.P. and Roelink, H. (1998) The teratogenic Veratrum alkaloid cyclopamine inhibits sonic hedgehog signal transduction. *Development*, **125**, 3553-3562.

Iseli, C., Jongeneel, C.V. and Bucher, P. (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 138-148.

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R., Tikunov, Y., Bovy, A., Chikate, Y., Singh, P., Rogachev, I., Beekwilder, J., Giri, A.P. and Aharoni, A. (2013) Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science*, **341**, 175-179.

Iven, T., Hornung, E., Heilmann, M. and Feussner, I. (2015) Synthesis of oleyl oleate wax esters in Arabidopsis thaliana and Camelina sativa seed oil. *Plant Biotechnol J*.

Izzi, L., Levesque, M., Morin, S., Laniel, D., Wilkes, B.C., Mille, F., Krauss, R.S., McMahon, A.P., Allen, B.L. and Charron, F. (2011) Boc and Gas1 each form distinct Shh receptor complexes with Ptch1 and are required for Shh-mediated cell proliferation. *Developmental cell*, **20**, 788-801.

Jiang, Y., Li, H., Li, P., Cai, Z. and Ye, W. (2005) Steroidal alkaloids from the bulbs of Fritillaria puqiensis. *Journal of natural products*, **68**, 264-267.

Jimeno, A., Weiss, G.J., Miller, W.H., Jr., Gettinger, S., Eigl, B.J., Chang, A.L., Dunbar, J., Devens, S., Faia, K., Skliris, G., Kutok, J., Lewis, K.D., Tibes, R., Sharfman, W.H., Ross, R.W. and Rudin, C.M. (2013) Phase I study of the Hedgehog pathway inhibitor IPI-926 in adult patients with solid tumors. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **19**, 2766-2774.

Johnson, M.T., Carpenter, E.J., Tian, Z., Bruskiewich, R., Burris, J.N., Carrigan, C.T., Chase, M.W., Clarke, N.D., Covshoff, S., Depamphilis, C.W., Edger, P.P., Goh, F., Graham, S., Greiner, S., Hibberd, J.M., Jordon-Thaden, I., Kutchan, T.M., Leebens-Mack, J., Melkonian, M., Miles, N., Myburg, H., Patterson, J., Pires, J.C., Ralph, P., Rolf, M., Sage, R.F., Soltis, D., Soltis, P., Stevenson, D., Stewart, C.N., Surek, B., Thomsen, C.J., Villarreal, J.C., Wu, X., Zhang, Y., Deyholos, M.K. and Wong, G.K. (2012) Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *PLoS One*, **7**, e50226.

Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bare, J.W., Bonifas, J.M., Quinn, A.G., Myers, R.M., Cox, D.R., Epstein, E.H., Jr. and Scott, M.P. (1996) Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science*, **272**, 1668-1671.

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W.E., Tuteja, R., Spillane, C., Robinson, S.J., Links, M.G., Clarke, C., Higgins, E.E., Huebert, T., Sharpe, A.G. and Parkin, I.A. (2014) The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. *Nature communications*, **5**, 3706.

Kaneko, K., Kawamura, N., Kuribayashi, T., Tanaka, M. and Mitsuhashi, H. (1978) Structures of two cevanine alkaloids, shinonomenine and veraflorizine, and a cevanidane alkaloid, procevine, isolated from illuminated veratrum. *Tetrahedron Letters*, **19**, 4801-4804.

Kaneko, K., Kawamura, N., Mitsuhash, H. and Ohsakik, K. (1979) Two new veratrum alkaloids, hosukinidine and epirubijervine from illuminated veratrum plant. *Chemical & pharmaceutical bulletin*, **27**, 2534-2536.

Kaneko, K., Mitsuhashi, H., Hirayama, K. and Ohmori, S. (1970a) 11-Deoxojervine as a precursor for jervine biosynthesis in Veratrum grandiflorum. *Phytochemistry*, **9**, 2497–2501.

Kaneko, K., Mitsuhashi, H., Hirayama, K. and Yoshida, N. (1970b) Biosynthesis of C-nor-D-homo-steroidal alkaloids from acetate-1-14C, cholesterol-4-14C and cholesterol-26-14C in Veratrum grandiflorum. *Phytochemistry*, **9**, 2489–2495.

Kaneko, K., Seto, H., Motoki, C. and Mitsuhashi, H. (1975) Biosynthesis of Rubijervine in Veratrum grandiflorum. *Phytochemistry*, **14**, 1295-1301.

Kaneko, K., Tanaka, H. and Mitsuhashi, H. (1977) Dormantinol, a possible precursor in solanidine biosynthesis, from budding Veratrum grandiflorum. *Phytochemistry*, **16**, 1247–1251.

Kaneko, K., Tanaka, M. and Mitsuhashi, H. (1976) Origin of nitrogen in the biosynthesis of solanidine by Veratrum grandiflorum. *Phytochemistry*, **15**, 1391–1393.

Kaneko, K., Watanabe, S., Taira, H. and Mitsuhashi, H. (1972) Conversion of solanidine to jerveratrum alkaloids in Veratrum grandiflorum. *Phytochemistry*, **11**, 3199-3202.

Karhadkar, S.S., Bova, G.S., Abdallah, N., Dhara, S., Gardner, D., Maitra, A., Isaacs, J.T., Berman, D.M. and Beachy, P.A. (2004) Hedgehog signalling in prostate regeneration, neoplasia and metastasis. *Nature*, **431**, 707-712.

Katoh, K., Asimenos, G. and Toh, H. (2009) Multiple alignment of DNA sequences with MAFFT. *Methods in molecular biology*, **537**, 39-64.

Katoh, K. and Standley, D.M. (2014) MAFFT: iterative refinement and additional methods. *Methods in molecular biology*, **1079**, 131-146.

Keeler, R.F. (1968) Teratogenic compounds of Veratrum californicum (Durand)-IV. First isolation of veratramine and alkaloid Q and a reliable method for isolation of cyclopamine. *Phytochemistry*, **7**, 303-306.

Keeler, R.F. (1969) Teratogenic compounds of veratrum californicum (durand) - VI: The structure of cyclopamine. *Phytochemistry*, **8**, 223-225.

Keeler, R.F. (1970a) Teratogenic compounds in Veratrum californicum (Durand) IX. Structure-activity relation. *Teratology*, **3**, 169-173.

Keeler, R.F. (1970b) Teratogenic compounds of Veratrum californicum (Durand) X. Cyclopia in rabbits produced by cyclopamine. *Teratology*, **3**, 175-180.

Keeler, R.F. (1975) Teratogenic effects of cyclopamine and jervine in rats, mice and hamsters. *Proc Soc Exp Biol Med*, **149**, 302-306.

Keeler, R.F. (1978a) Cyclopamine and related steroidal alkaloid teratogens: their occurrence, structural relationship, and biologic effects. *Lipids*, **13**, 708-715.

Keeler, R.F. (1978b) Reducing Incidence of Plant-Caused Congenital Deformities in Livestock by Grazing Management. *Journal of Range Management*, **31**, 355-360.

Keeler, R.F. and Binns, W. (1966a) Teratogenic compounds of Veratrum californicum (Durand). I. Preparation and characterization of fractions and alkaloids for biologic testing. *Canadian journal of biochemistry*, **44**, 819-828.

Keeler, R.F. and Binns, W. (1966b) Teratogenic compounds of Veratrum californicum (Durand). II. Production of ovine fetal cyclopia by fractions and alkaloid preparations. *Canadian journal of biochemistry*, **44**, 829-838.

Keeler, R.F. and Binns, W. (1968) Teratogenic compounds in Veratrum californicum (Durand) V. Comparison of cyclopian effects of steroidal alkaloids from the plant and structurally related compounds from other sources. *Teratology*, **1**, 5-10.

Keeler, R.F., Young, S. and Brown, D. (1976) Spina bifida, exencephaly, and cranial bleb produced in hamsters by the solanum alkaloid solasodine. *Res Commun Chem Pathol Pharmacol.* , **13**, 723-730.

Kilgore, M.B., Augustin, M.M., Starks, C.M., O'Neil-Johnson, M., May, G.D., Crow, J.A. and Kutchan, T.M. (2014) Cloning and characterization of a norbelladine 4'-O-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in Narcissus sp. aff. pseudonarcissus. *PLoS One*, **9**, e103223.

Kim, H.J., Silva, J.E., Vu, H.S., Mockaitis, K., Nam, J.W. and Cahoon, E.B. (2015) Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in Camelina seeds. *Journal of experimental botany*, **66**, 4251-4265.

Ko, A. (2015) FOLFIRINOX Plus IPI-926 for Advanced Pancreatic Adenocarcinoma. In *2000-2015Nov02* (ClinicalTrials.gov ed. Bethesda (MD): National Library of Medicine (US): https://clinicaltrials.gov/ct2/show/NCT01383538.

Kristensen, C., Morant, M., Olsen, C.E., Ekstrom, C.T., Galbraith, D.W., Moller, B.L. and Bak, S. (2005) Metabolic engineering of dhurrin in transgenic Arabidopsis plants with marginal inadvertent effects on the metabolome and transcriptome. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 1779-1784.

Kumar, S.K., Roy, I., Anchoori, R.K., Fazli, S., Maitra, A., Beachy, P.A. and Khan, S.R. (2008) Targeted inhibition of hedgehog signaling by cyclopamine prodrugs for advanced prostate cancer. *Bioorganic & medicinal chemistry*, **16**, 2764-2768.

Kusano, G., Takahashi, A., Sugiyama, K. and Nozoe, S. (1987) Antifungal properties of solanum alkaloids. *Chemical & pharmaceutical bulletin*, **35**, 4862-4867.

Lee, S.T., Welch, K.D., Panter, K.E., Gardner, D.R., Garrossian, M. and Chang, C.W. (2014) Cyclopamine: from cyclops lambs to cancer treatment. *Journal of agricultural and food chemistry*, **62**, 7355-7362.

Lee, S.T., Wong, P.F., Cheah, S.C. and Mustafa, M.R. (2011) Alpha-tomatine induces apoptosis and inhibits nuclear factor-kappa B activation on human prostatic adenocarcinoma PC-3 cells. *PLoS One*, **6**, e18915.

Lee, S.T., Wong, P.F., He, H., Hooper, J.D. and Mustafa, M.R. (2013a) Alpha-tomatine attenuation of in vivo growth of subcutaneous and orthotopic xenograft tumors of human

prostate carcinoma PC-3 cells is accompanied by inactivation of nuclear factor-kappa B signaling. *PLoS One*, **8**, e57708.

Lee, S.T., Wong, P.F., Hooper, J.D. and Mustafa, M.R. (2013b) Alpha-tomatine synergises with paclitaxel to enhance apoptosis of androgen-independent human prostate cancer PC-3 cells in vitro and in vivo. *Phytomedicine*, **20**, 1297-1305.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., Debono, A., Durrett, T.P., Franke, R.B., Graham, I.A., Katayama, K., Kelly, A.A., Larson, T., Markham, J.E., Miquel, M., Molina, I., Nishida, I., Rowland, O., Samuels, L., Schmid, K.M., Wada, H., Welti, R., Xu, C., Zallot, R. and Ohlrogge, J. (2013) Acyl-lipid metabolism. *The Arabidopsis book / American Society of Plant Biologists*, **11**, e0161.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

Li, H.J., Jiang, Y. and Li, P. (2006) Chemistry, bioactivity and geographical diversity of steroidal alkaloids from the Liliaceae family. *Nat Prod Rep*, **23**, 735-752.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J. and Wang, J. (2010) De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*, **20**, 265-272.

Lin, T.L. and Matsui, W. (2012) Hedgehog pathway as a drug target: Smoothened inhibitors in development. *Onco Targets Ther*, **5**, 47-58.

Lin, T.L., Wang, Q.H., Brown, P., Peacock, C., Merchant, A.A., Brennan, S., Jones, E., McGovern, K., Watkins, D.N., Sakamoto, K.M. and Matsui, W. (2010) Self-renewal of acute lymphocytic leukemia cells is limited by the Hedgehog pathway inhibitors cyclopamine and IPI-926. *PLoS One*, **5**, e15262.

Linsmaier E.M., S.F. (1965) Organic growth factor requirements of tobacco tissue cultures. *Physiologia Plantarum*, **18**, 100-127.

Liu, J., Rice, A., McGlew, K., Shaw, V., Park, H., Clemente, T., Pollard, M., Ohlrogge, J. and Durrett, T.P. (2015) Metabolic engineering of oilseed crops to produce high levels of novel acetyl glyceride oils with reduced viscosity, freezing point and calorific value. *Plant Biotechnol J*, **13**, 858-865.

Lottaz, C., Iseli, C., Jongeneel, C.V. and Bucher, P. (2003) Modeling sequencing errors by combining Hidden Markov models. *Bioinformatics*, **19 Suppl 2**, ii103-112.

Lu, C. and Kang, J. (2008) Generation of transgenic plants of a potential oilseed crop Camelina sativa by Agrobacterium-mediated transformation. *Plant cell reports*, **27**, 273-278.

Ma, H., Li, H.Q. and Zhang, X. (2013) Cyclopamine, a naturally occurring alkaloid, and its analogues may find wide applications in cancer therapy. *Curr Top Med Chem*, **13**, 2208-2215.

Malik, M.R., Yang, W., Patterson, N., Tang, J., Wellinghoff, R.L., Preuss, M.L., Burkitt, C., Sharma, N., Ji, Y., Jez, J.M., Peoples, O.P., Jaworski, J.G., Cahoon, E.B. and Snell, K.D. (2015) Production of high levels of poly-3-hydroxybutyrate in plastids of Camelina sativa seeds. *Plant Biotechnol J*, **13**, 675-688.

Mansour, M.P., Shrestha, P., Belide, S., Petrie, J.R., Nichols, P.D. and Singh, S.P. (2014) Characterization of oilseed lipids from "DHA-producing Camelina sativa": a new transformed land plant containing long-chain omega-3 oils. *Nutrients*, **6**, 776-789.

Masamune, T., Mori, Y., Takagui, E.M. and Murai, A. (1964) A new alkaloid from Veratrum species, 11-deoxojervine. *Biological & pharmaceutical bulletin*, **5**, 913-917.

McMillan, R. and Matsui, W. (2012) Molecular pathways: the hedgehog signaling pathway in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **18**, 4883-4888.

Michael, T.P., Breton, G., Hazen, S.P., Priest, H., Mockler, T.C., Kay, S.A. and Chory, J. (2008) A morning-specific phytohormone gene expression program underlying rhythmic plant growth. *PLoS Biol*, **6**, e225.

Mishra, B.B. and Tiwari, V.K. (2011) Natural products: an evolving role in future drug discovery. *European journal of medicinal chemistry*, **46**, 4769-4807.

Mithofer, A. and Boland, W. (2012) Plant defense against herbivores: chemical aspects. *Annual review of plant biology*, **63**, 431-450.

Mizutani, M. (2012) Impacts of diversification of cytochrome P450 on plant metabolism. *Biological & pharmaceutical bulletin*, **35**, 824-832.

Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., McEntee, C., Kay, S.A. and Chory, J. (2007) The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb Symp Quant Biol*, **72**, 353-363.

Murashige, T.S., F. (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum*, **15**, 473–497.

Nachtergaele, S., Mydock, L.K., Krishnan, K., Rammohan, J., Schlesinger, P.H., Covey, D.F. and Rohatgi, R. (2012) Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nature chemical biology*, **8**, 211-220.

Nakamura, K., Sasajima, J., Mizukami, Y., Sugiyama, Y., Yamazaki, M., Fujii, R., Kawamoto, T., Koizumi, K., Sato, K., Fujiya, M., Sasaki, K., Tanno, S., Okumura, T., Shimizu, N., Kawabe, J., Karasaki, H., Kono, T., Ii, M., Bardeesy, N., Chung, D.C. and Kohgo, Y. (2010) Hedgehog promotes neovascularization in pancreatic cancers by regulating Ang-1 and IGF-1 expression in bone-marrow derived pro-angiogenic cells. *PLoS One*, **5**, e8824.

Nedelcu, D., Liu, J., Xu, Y., Jao, C. and Salic, A. (2013) Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nature chemical biology*, **9**, 557-564.

Niewiadomski, P., Kong, J.H., Ahrends, R., Ma, Y., Humke, E.W., Khan, S., Teruel, M.N., Novitch, B.G. and Rohatgi, R. (2014) Gli protein activity is controlled by multisite phosphorylation in vertebrate Hedgehog signaling. *Cell reports*, **6**, 168-181.

Nims, E., Dubois, C.P., Roberts, S.C. and Walker, E.L. (2006) Expression profiling of genes involved in paclitaxel biosynthesis for targeted metabolic engineering. *Metabolic engineering*, **8**, 385-394.

Ohyama, K., Suzuki, M., Kikuchi, J., Saito, K. and Muranaka, T. (2009) Dual biosynthetic pathways to phytosterol via cycloartenol and lanosterol in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 725-730.

Olive, K.P., Jacobetz, M.A., Davidson, C.J., Gopinathan, A., McIntyre, D., Honess, D., Madhu, B., Goldgraben, M.A., Caldwell, M.E., Allard, D., Frese, K.K., Denicola, G., Feig, C., Combs, C., Winter, S.P., Ireland-Zecchini, H., Reichelt, S., Howat, W.J., Chang, A., Dhara, M., Wang, L., Ruckert, F., Grutzmann, R., Pilarsky, C., Izeradjene, K., Hingorani, S.R., Huang, P., Davies, S.E., Plunkett, W., Egorin, M., Hruban, R.H., Whitebread, N., McGovern, K., Adams, J., Iacobuzio-Donahue, C., Griffiths, J. and Tuveson, D.A. (2009) Inhibition of Hedgehog signaling enhances delivery of chemotherapy in a mouse model of pancreatic cancer. *Science*, **324**, 1457-1461.

Onrubia, M., Moyano, E., Bonfill, M., Palazon, J., Goossens, A. and Cusido, R.M. (2011) The relationship between TXS, DBAT, BAPT and DBTNBT gene expression and taxane production during the development of Taxus baccata plantlets. *Plant science : an international journal of experimental plant biology*, **181**, 282-287.

Peacock, C.D., Wang, Q., Gesell, G.S., Corcoran-Schwartz, I.M., Jones, E., Kim, J., Devereux, W.L., Rhodes, J.T., Huff, C.A., Beachy, P.A., Watkins, D.N. and Matsui, W. (2007) Hedgehog signaling maintains a tumor stem cell compartment in multiple

myeloma. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 4048-4053.

Peluso, M.O., Campbell, V.T., Harari, J.A., Tibbitts, T.T., Proctor, J.L., Whitebread, N., Conley, J.M., White, K.F., Kutok, J.L., Read, M.A., McGovern, K. and Faia, K.L. (2014) Impact of the Smoothened inhibitor, IPI-926, on smoothened ciliary localization and Hedgehog pathway activity. *PLoS One*, **9**, e90534.

Petrie, J.R., Shrestha, P., Belide, S., Kennedy, Y., Lester, G., Liu, Q., Divi, U.K., Mulder, R.J., Mansour, M.P., Nichols, P.D. and Singh, S.P. (2014) Metabolic engineering Camelina sativa with fish oil-like levels of DHA. *PLoS One*, **9**, e85061.

Porter, J.A., Young, K.E. and Beachy, P.A. (1996) Cholesterol modification of hedgehog signaling proteins in animal development. *Science*, **274**, 255-259.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A. and Finn, R.D. (2012) The Pfam protein families database. *Nucleic acids research*, **40**, D290-301.

Putnam, D.H., Budin, J.T., Field, L.A. and M., B.W. (1993) Camelina: A Promising Low-Input Oilseed. In *New Crops* (Janick, J. and Simon, J.E. eds). New York: Wiley.

Renoux, B., Legigan, T., Bensalma, S., Chadeneau, C., Muller, J.M. and Papot, S. (2011) A new cyclopamine glucuronide prodrug with improved kinetics of drug release. *Org Biomol Chem*, **9**, 8459-8464.

Rohatgi, R. and Scott, M.P. (2007) Patching the gaps in Hedgehog signalling. *Nat Cell Biol*, **9**, 1005-1009.

Romer, J.T., Kimura, H., Magdaleno, S., Sasai, K., Fuller, C., Baines, H., Connelly, M., Stewart, C.F., Gould, S., Rubin, L.L. and Curran, T. (2004) Suppression of the Shh pathway using a small molecule inhibitor eliminates medulloblastoma in Ptc1(+/-)p53(-/-) mice. *Cancer Cell*, **6**, 229-240.

Rosco, A., Pauli, H.H., Priesner, W. and Kutchan, T.M. (1997) Cloning and heterologous expression of NADPH-cytochrome P450 reductases from the Papaveraceae. *Arch Biochem Biophys*, **348**, 369-377.

Ruiz-Lopez, N., Haslam, R.P., Napier, J.A. and Sayanova, O. (2014) Successful high-level accumulation of fish oil omega-3 long-chain polyunsaturated fatty acids in a transgenic oilseed crop. *The Plant journal : for cell and molecular biology*, **77**, 198-208.

Saito, K., Hirai, M.Y. and Yonekura-Sakakibara, K. (2008) Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. *Trends in plant science*, **13**, 36-43.

Sharpe, H.J., Wang, W., Hannoush, R.N. and de Sauvage, F.J. (2015) Regulation of the oncoprotein Smoothened by small molecules. *Nature chemical biology*, **11**, 246-255.

Shi, Y., Moura, U., Opitz, I., Soltermann, A., Rehrauer, H., Thies, S., Weder, W., Stahel, R.A. and Felley-Bosco, E. (2012) Role of hedgehog signaling in malignant pleural mesothelioma. *Clinical cancer research : an official journal of the American Association for Cancer Research*, **18**, 4646-4656.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J. and Birol, I. (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res*, **19**, 1117-1123.

Song, J., Naylor-Adelberg, J., White, S., Mann, D. and Adelberg, J. (2014) Establishing clones of Veratrum californicum, a native medicinal species, for micropropagation. *In Vitro Cellular & Developmental Biology - Plant*, **50**, 337-344.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688-2690.

Stecca, B., Mas, C., Clement, V., Zbinden, M., Correa, R., Piguet, V., Beermann, F. and Ruiz, I.A.A. (2007) Melanomas require HEDGEHOG-GLI signaling regulated by

interactions between GLI1 and the RAS-MEK/AKT pathways. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 5895-5900.

Taipale, J., Chen, J.K., Cooper, M.K., Wang, B., Mann, R.K., Milenkovic, L., Scott, M.P. and Beachy, P.A. (2000) Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature*, **406**, 1005-1009.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*, **28**, 2731-2739.

Thayer, S.P., di Magliano, M.P., Heiser, P.W., Nielsen, C.M., Roberts, D.J., Lauwers, G.Y., Qi, Y.P., Gysin, S., Fernandez-del Castillo, C., Yajnik, V., Antoniu, B., McMahon, M., Warshaw, A.L. and Hebrok, M. (2003) Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature*, **425**, 851-856.

Tremblay, M.R., Lescarbeau, A., Grogan, M.J., Tan, E., Lin, G., Austad, B.C., Yu, L.C., Behnke, M.L., Nair, S.J., Hagel, M., White, K., Conley, J., Manna, J.D., Alvarez-Diez, T.M., Hoyt, J., Woodward, C.N., Sydor, J.R., Pink, M., MacDougall, J., Campbell, M.J., Cushing, J., Ferguson, J., Curtis, M.S., McGovern, K., Read, M.A., Palombella, V.J., Adams, J. and Castro, A.C. (2009) Discovery of a potent and orally active hedgehog pathway antagonist (IPI-926). *Journal of medicinal chemistry*, **52**, 4400-4418.

Tschesche, R. and Brennecke, H. (1980) Side chain functionalization of cholesterol in the biosynthesis of solasodine in Solanum laciniatum *Phytochemistry*, **19**, 1449–1451.

Turano, F.J. and Fang, T.K. (1998) Characterization of two glutamate decarboxylase cDNA clones from Arabidopsis. *Plant physiology*, **117**, 1411-1421.

UniProt, C. (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic acids research*, **41**, D43-47.

Unterlinner, B., Lenz, R. and Kutchan, T.M. (1999) Molecular cloning and functional expression of codeinone reductase: the penultimate enzyme in morphine biosynthesis in the opium poppy Papaver somniferum. *The Plant journal : for cell and molecular biology*, **18**, 465-475.

Varjosalo, M. and Taipale, J. (2007) Hedgehog signaling. *J Cell Sci*, **120**, 3-6.

Varnat, F., Duquet, A., Malerba, M., Zbinden, M., Mas, C., Gervaz, P. and Ruiz i Altaba, A. (2009) Human colon cancer epithelial cells harbour active HEDGEHOG-GLI signalling that is essential for tumour growth, recurrence, metastasis and stem cell survival and expansion. *EMBO molecular medicine*, **1**, 338-351.

Wall, P.K., Leebens-Mack, J., Muller, K.F., Field, D., Altman, N.S. and dePamphilis, C.W. (2008) PlantTribes: a gene and gene family resource for comparative genomics in plants. *Nucleic acids research*, **36**, D970-976.

Watkins, D.N., Berman, D.M., Burkholder, S.G., Wang, B., Beachy, P.A. and Baylin, S.B. (2003) Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature*, **422**, 313-317.

Wechsler-Reya, R. and Scott, M.P. (2001) The developmental biology of brain tumors. *Annu Rev Neurosci*, **24**, 385-428.

Weid, M., Ziegler, J. and Kutchan, T.M. (2004) The roles of latex and the vascular bundle in morphine biosynthesis in the opium poppy, Papaver somniferum. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 13957-13962.

Yauch, R.L., Gould, S.E., Scales, S.J., Tang, T., Tian, H., Ahn, C.P., Marshall, D., Fu, L., Januario, T., Kallop, D., Nannini-Pepe, M., Kotkow, K., Marsters, J.C., Rubin, L.L. and de Sauvage, F.J. (2008) A paracrine requirement for hedgehog signalling in cancer. *Nature*, **455**, 406-410.

Zhang, J., Garrossian, M., Gardner, D., Garrossian, A., Chang, Y.T., Kim, Y.K. and Chang, C.W. (2008) Synthesis and anticancer activity studies of cyclopamine derivatives. *Bioorganic & medicinal chemistry letters*, **18**, 1359-1363.

Zhao, C., Chen, A., Jamieson, C.H., Fereshteh, M., Abrahamsson, A., Blum, J., Kwon, H.Y., Kim, J., Chute, J.P., Rizzieri, D., Munchhof, M., VanArsdale, T., Beachy, P.A. and Reya, T. (2009) Hedgehog signalling is essential for maintenance of cancer stem cells in myeloid leukaemia. *Nature*, **458**, 776-779.

Zhou, C.L., J.; Ye, W.; Liu, C.; Tan, R. (2003) Neoverataline A and B, two antifungal alkaloids with a novel carbon skeleton from *Veratrum taliense*. *Tetrahedron*, **59**, 5743-5747.

Ziegler, J., Voigtlander, S., Schmidt, J., Kramell, R., Miersch, O., Ammer, C., Gesell, A. and Kutchan, T.M. (2006) Comparative transcript and alkaloid profiling in Papaver species identifies a short chain dehydrogenase/reductase involved in morphine biosynthesis. *The Plant journal : for cell and molecular biology*, **48**, 177-192.